WORLD POLITICS

A Quarterly Journal of International Relations

Volume 52, Number 1   October 1999

The editors invite submission of articles (up to 10,000 words, including notes) bearing upon problems in international relations and comparative politics. Manuscripts and notes should be double-spaced throughout and submitted in triplicate, together with an abstract. Notes should be placed at the end of the manuscript. Address: WORLD POLITICS, Bendheim Hall, Princeton University, Princeton, NJ 08544.

Authors may expect to be notified of the editors' decision within four months. Procedures for reviewing manuscripts are based on the anonymity of the author and the confidentiality of readers' and editors' reports; hence, self-references should be removed. Manuscripts are refereed by the editors and by other scholars at Princeton and at other institutions. WORLD POLITICS does not accept manuscripts that have already been published, are scheduled for publication elsewhere, or have been simultaneously submitted to another journal; this applies to both print and online formats. Statements of fact and opinion appearing in the journal are made on the responsibility of the authors alone and do not imply the endorsement of the editors or publisher. The journal does not publish communications to the editor or rejoinders to specific articles; scholars who believe they have been challenged are encouraged to submit an article that will advance the scholarly debate. The WORLD POLITICS home page is found at http://www.wws.princeton.edu/world_politics.

Reprint permission. Permissions Department, The Johns Hopkins University Press, 2715 N. Charles Street, Baltimore, MD 21218-4319.

Subscriptions. WORLD POLITICS, The Johns Hopkins University Press, Journals Publishing Division, 2715 N. Charles Street, Baltimore, MD 21218-4319. Phone: (410) 516-6987; FAX: (410) 516-6968; toll-free (800) 548-1784. Individuals: $28 per year; institutions: $86 per year. Single copies: individuals, $8.00; institutions, $21.00. For postage and handling in Canada and Mexico, add $4.00 per year; outside of North America, $9.00.

Back Issues. WORLD POLITICS is available in microfilm or xerographic copies from University Microfilms, Ann Arbor, MI 48106. Two years after publication, complete volumes are available in microfiche form from KTO Microform, Route 100, Millwood, NY 10546. Individual issues in the current and two preceding volumes may be ordered from The Johns Hopkins University Press, Journals Publishing Division, 2715 N Charles Street, Baltimore, MD 21218-4319.

Advertising. Journals Advertising Coordinator, The Johns Hopkins University Press, Journals Publishing Division, 2715 N. Charles Street, Baltimore, MD 21218-4319.

WORLD POLITICS is indexed by *ABC POL SCI, Book Review Index, Combined Retrospective Index Sets (CRIS), Combined Retrospective Index to Book Reviews in Scholarly Journals, 1886–1974 (RSJ), PAIS Bulletin Social Science Index,* and *Periodica Islamica;* it is abstracted and indexed in *United States Political Science Documents, International Bibliography of Periodical Literature (IBZ), International Bibliography of Book Reviews (IBR),* and *Current Military and Political Literature.* Abstracts of articles also appear in *Book Review Digest, Historical Abstracts, International Development Abstracts, International Political Science Abstracts, Political Science Abstracts,* and *Social Sciences Abstracts.*

WORLD POLITICS (ISSN 0043-8871). Published quarterly by The Johns Hopkins University Press. Vol. 52, No. 1, October 1999. Periodicals postage paid at Baltimore, MD, and additional mailing offices. POSTMASTER: Send address changes to WORLD POLITICS, The Johns Hopkins University Press, Journals Publishing Division, 2715 N. Charles Street, Baltimore, MD 21218-4319. Printed in the United States of America by Capital City Press.

WORLD POLITICS

.52 • October 1999 No. 1

CONTENTS

THE CONTRIBUTORS

JOHN R. ONEAL is Professor of Political Science at the University of Alabama at Tuscaloosa, and spend spring term 2000 as a Fulbright Scholar and Fellow at the Norwegian Nobel Institute in C He is the author (with Bruce Russett) of *Triangulating Peace: Democracy, Interdependence, and Inte tional Organizations* (forthcoming).

BRUCE RUSSETT is Dean Acheson Professor of Political Science and Director of United Nations St ies at Yale University. He is the author (with John R. Oneal) of *Triangulating Peace: Democracy, In dependence, and International Organizations* (forthcoming).

STROM C. THACKER is Assistant Professor of International Relations at Boston University. He is t author of *Big Business, the State, and Free Trade: Constructing Coalitions in Mexico* (forthcoming). He currently conducting research on the polities of bilateral and multilateral aid and on neoliberalism a democracy in Latin America.

DEBORAH J. YASHAR is Assistant Professor of Politics and International Affairs at Princeton Unive sity. She is the author of *Demanding Democracy: Reform and Reaction in Costa Rica and Guatemal 1870s–1990s* (1997) and of articles and chapters on political regimes, social movements, and ethn politics. She is currently writing a book that examines the intersection of democracy, citizenship, an indigenous movements in Latin America.

KATRINA BURGESS is Assistant Professor of Political Science in the Maxwell School of Citizenshi and Public Affairs at Syracuse University. She has published several articles on union politics in Spai and is coeditor (with Abraham F. Lowenthal) of *The California-Mexico Connection* (1993). Currentl she is completing a book manuscript on party-union relations in Mexico, Spain, and Venezuela and i beginning a new comparative project on the adaptive capacities of mass-based parties faced with crise of representation.

WORLD POLITICS

ol. 52 • January 2000 • No. 2

CONTENTS

THE CONTRIBUTORS

M. VICTORIA MURILLO is Assistant Professor of Political Science at Yale University. She is t author of *Divided Loyalties: Labor Unions and Market Reforms in Latin America* (forthcomin She has also contributed to numerous academic journals and edited volumes on themes such teachers' unions and the decentralization of education in Argentina and Mexico, and mark oriented reforms and Argentine corporatism.

EVA BELLIN is Associate Professor of Government at Harvard University. She is the author *Stalled Democracy: Capitalist Industrialization and the Paradox of State Sponsorship in Tunisia, t Middle East, and Beyond* (forthcoming).

DANIEL PHILPOTT is Assistant Professor of Political Science at the University of Californi Santa Barbara. He is the author of *Revolutions in Sovereignty* (forthcoming) and has publish articles on sovereignty and self-determination.

RICHARD W. WILSON is Professor of Political Science at Rutgers-The State University of Ne Jersey in New Brunswick. Among other books, he has published *Labyrinth: An Essay on the P litical Psychology of Change* (1988) and *Compliance Ideologies: Rethinking Political Culture* (199 His work of the past two decades emphasized the development of theory in political culture; h current work focuses on the usefulness of biological models for understanding political cultur

WORLD POLITICS

Vol. 52 • April 2000 • No. 3

CONTENTS

THE CONTRIBUTORS

CONSUELO CRUZ is Assistant Professor of Political Science at Columbia University. She ha published several articles on Latin American politics. She is also completing a book manuscrip "World-Making in the Tropics: The Politics of Fate and Possibility," which examines the role c political culture in shaping the contrasting development paths of Nicaragua and Costa Rica.

TORBEN IVERSEN is Associate Professor of Government at Harvard University. He is the autho of *Contested Economic Institutions: The Politics of Macroeconomics and Wage Bargaining* (1999) an coeditor of *Unions, Employers and Central Bankers: Macroeconomic Coordination and Institution Change in Social Market Economies* (2000), and he is working on a book exploring the intersectio between social protection and the nature of skills.

THOMAS R. CUSACK is Senior Research Fellow at the Wissenschaftszentrum Berlin für Sozial forschung. He is the coauthor of *Exploring Realpolitik* (1990) and coeditor of *The Process of Wa* (1995), and he is completing a book on institutions, political culture, and local government per formance in Germany.

DAVID RUEDA is a Ph.D. candidate in the Department of Government, Cornell University. H is completing a dissertation on the politics of unemployment and the effects of partisan differ ences over economic policy in OECD countries.

JONAS PONTUSSON is Associate Professor of Government at Cornell University. His publication include *The Limits of Social Democracy: Investment Politics in Sweden* (1992) and, as coeditor, *Bar gaining for Change: Union Politics and Intra-Class Conflict in Western Europe and North Americ* (1992) and *Unions, Employers and Central Banks: Wage Bargaining and Macroeconomic Regimes i an Integrating Europe* (2000). His current research focuses on the interface of social polic regimes and labor markets.

RENSKE DOORENSPLEET is a Ph.D. candidate in the Political Science Department at Leide University, the Netherlands. Her current research considers the conditions and challenges o democratic transitions around the world since the late 1980s.

JONATHAN KIRSHNER is Associate Professor of Government at Cornell University. He is the au thor of *Currency and Coercion: The Political Economy of International Monetary Power* (1995) an is currently working on a book about how financial interests shape and constrain grand strateg

WORLD POLITICS

Vol. 52 • July 2000 • No. 4

CONTENTS

THE CONTRIBUTORS

NICHOLAS SAMBANIS is Economist in the Development Economics Research Group at the World Bank and Lecturer in Political Science at Yale University. He researches the political economy of civil wars.

PATRICK HELLER is an Associate Professor of Sociology at Brown University. He is the author of *The Labor of Development: Workers and the Transformation of Capitalism in Kerala, India* (1999) and is currently conducting research on democratization and local governance in South Africa.

HILARY APPEL is Assistant Professor of Government at Claremont McKenna College. She is the author of several articles on the political-economic transformation in Russia and Eastern Europe. She is currently completing a book manuscript on the ideological foundations of post-communist economic reform.

RICHARD NED LEBOW is Director of the Mershon Center and Professor of Political Science, History and Psychology at The Ohio State University. His recent books include *We All Lost the Cold War* (1994), coauthored with Janice Gross Stein, and *The Art of Bargaining* (1996). He has a forthcoming study, *Ethics, War and Society*, a novel, *Play it Again Ilsa*, and three coedited works: *Unmaking the West: Counterfactual and Contingency*, *Learning from the Cold War*, and *Theory and Evidence in Comparative Politics and International Relations*.

SHU-YUN MA is Associate Professor of Government and Public Administration at the Chinese University of Hong Kong. He is the author of a number of recent articles that have appeared in journals of comparative politics and communist and postcommunist studies..

ABSTRACTS

THE KANTIAN PEACE

THE PACIFIC BENEFITS OF DEMOCRACY, INTERDEPENDENCE, AND INTERNATIONAL ORGANIZATIONS, 1885–1992

By JOHN R. ONEAL and BRUCE RUSSETT

The authors test Kantian and realist theories of interstate conflict using data extending over more than a century, treating those theories as complementary rather than competing. As the classical liberals believed, democracy, economic interdependence, and international organizations have strong and statistically significant effects on reducing the probability that states will be involved in militarized disputes. Moreover, the benefits are not limited to the cold war era. Some realist influences, notably distance and power predominance, also reduce the likelihood of interstate conflict. The character of the international system, too, affects the probability of dyadic disputes. The consequences of having a strong hegemonic power vary, but high levels of democracy and interdependence in the international system reduce the probability of conflict for all dyads, not just for those that are democratic or dependent on trade.

THE HIGH POLITICS OF IMF LENDING

By STROM C. THACKER

Analysts have long suspected that politics affects the lending patterns of the International Monetary Fund (IMF), but none have adequately specified or systematically tested competing explanations. This paper develops a political explanation of IMF lending and tests it statistically on the developing countries between 1985 and 1994. It finds that political realignment toward the United States, the largest power in the IMF, increases a country's probability of receiving an IMF loan. A country's static political alignment position has no significant impact during this period, suggesting that these processes are best modeled dynamically. An analysis of two subsamples rejects the hypothesis that the IMF has become less politicized since the end of the cold war and suggests that the influence of politics has actually increased since 1990. The behavior of multilateral organizations is still driven by the political interests of their more powerful member states.

DEMOCRACY, INDIGENOUS MOVEMENTS, AND THE POSTLIBERAL CHALLENGE IN LATIN AMERICA

By DEBORAH J. YASHAR

Scholars of democratic consolidation have come to focus on the links between political institutions and enduring regime outcomes. This article takes issue with the conceptual and analytical underpinnings of this literature by highlighting how new political institutions, rather than securing democratic politics, have in fact had a more checkered effect. It delineates why the theoretical expectations of the democratic consolidation literature have not been realized and draws, by example, on the contemporary ethnic movements that are now challenging third-wave democracies. In particular, it highlights how contemporary indigenous movements, emerging in response to unevenly institutionalized reforms, pose a postliberal challenge to Latin America's newly founded democracies. These movements have sparked political debates and constitutional reforms over community rights, territorial autonomy, and a multiethnic citizenry. As a whole, they have laid bare the weakness of state institutions, the contested terms of democracy, and the indeterminacy of ethnic accommodation in the region. As such, these movements highlight the need to qualify somewhat premature and narrow discussions of democratic consolidation in favor of a broader research agenda on democratic politics.

Loyalty Dilemmas and Market Reform

PARTY-UNION ALLIANCES UNDER STRESS IN MEXICO, SPAIN, AND VENEZUELA

By KATRINA BURGESS

Market reform has dealt a serious blow to traditional alliances between governing parties and labor unions. This article examines the fate of these alliances by applying a revised version of Albert Hirschman's schema of exit, voice, and loyalty to party-union relations in Mexico, Spain and Venezuela. After refining the concept of loyalty, the author argues that it is embedded in the principles and norms on which these alliances are based. Market reform places party-affiliated labor leaders in a "loyalty dilemma" in which they have no choice but to behave disloyally toward one set of claimants. Their propensity to respond with either voice or exit depends on their vulnerability to reprisals for disloyal behavior and the party's capacity to retain their loyalty even in the face of sacrifices imposed on workers and unions. Both variables are linked to the authority structures in which labor and party leaders find themselves. In the short to medium run the alliances most likely to survive are those in which labor leaders have significant autonomy from their bases and/or in which the party is able and willing to challenge its own executive. In the long run, however, even these alliances may be vulnerable to collapse because of popular frustrations with the inadequacy of interest representation and the multiple pressures on political organizations to adapt to a more fluid and uncertain environment.

THE KANTIAN PEACE
The Pacific Benefits of Democracy, Interdependence, and International Organizations, 1885–1992

By JOHN R. ONEAL and BRUCE RUSSETT*

JUST over two hundred years ago Immanuel Kant suggested that "republican constitutions," a "commercial spirit" of international trade, and a federation of interdependent republics would provide the basis for perpetual peace. The alternative, even starker in the nuclear era than in 1795, would be peace of a different sort: "a vast grave where all the horrors of violence and those responsible for them would be buried."[1] Consequently, Kant declared, we have a duty to work for peaceful international relations. Though he emphasized the absolute character of this moral imperative, he was no idealist; rather, he believed that natural processes based on self-interest impelled individuals to act in ways that would eventually produce a lasting and just peace. Kant was also realistic. He acknowledged that war was inherent in the anarchic international system and therefore cautioned that nations must act prudently until the federation of interdependent republics was established. But he also knew that the mechanisms of power politics produce only temporary respite from conflict, not lasting solutions.

Over the past half century much of the world has been at peace. Understanding that phenomenon, its causes and trajectory, is the fundamental challenge for international relations scholars today. We seek to show that Kant's realistic statement of liberal theory provides useful guidance for this task. Most political scientists now agree that the contemporary peacefulness can be traced in part to the so-called democratic peace, wherein established democratic states have fought no

* We thank the Carnegie Corporation of New York, the Ford Foundation, and the National Science Foundation for financial support; Zeev Maoz for comments; and Jennifer Beam, Margit Bussmann, Soo Yeon Kim, Yury Omelchenko, Brian Radigan, and Jacob Sullivan for data collection and management.

[1] Kant, *Perpetual Peace: A Philosophical Sketch*, in *Kant's Political Writings*, ed. Hans Reiss (Cambridge: Cambridge University Press, 1970), 105. See also James Bohman and Matthias Lutz-Bachmann, eds., *Perpetual Peace: Essays on Kant's Cosmopolitan Ideal* (Cambridge: MIT Press, 1997).

international wars with one another and the use or threat of force among them, even at low levels, has been rare.[2] This view is incomplete, however, because it fails to recognize the pacific benefits of the other liberal elements of Kant's program for peace. Moreover, the term hides the vigorous theoretical controversy about the processes underlying this separate peace—over whether democracy is really even its cause and over the degree to which the empirical phenomenon existed in other eras.

These theoretical and empirical concerns are linked. If, for example, peaceful relations among democracies during the cold war era were simply a consequence of their shared security interests vis-à-vis the opposing alliance system in a bipolar world, then their peacefulness would be spuriously related to the character of their regimes. The same conclusion would result if the democratic peace could be attributed to the hegemonic power of the United States to suppress conflict among its allies or to East-West differences in preferences unrelated to underlying differences in regimes.[3] One would not then expect to find a separate peace among democratic states in other periods evincing different patterns of interstate relations. We address these questions by reporting analyses covering 1885–1992, to show that peaceful relations among democracies existed throughout the twentieth century.[4] Extending the historical domain also allows us to assess the effect of the changing character of the international system on interstate relations.[5]

[2] By convention in the social science literature, war is defined as a conflict between two recognized sovereign members of the international system that results in at least one thousand battle deaths. The most complete data on militarized international disputes (MIDs), compiled by Stuart Bremer and his colleagues, are available at http://pss.la.psu.edu/MID_DATA.HTM. The democracy data we employ were compiled by Keith Jaggers and Ted Robert Gurr, "Tracking Democracy's Third Wave with the Polity III Data," *Journal of Peace Research* 32, no. 4 (1995), available at http://isere.colorado.edu-/pub/datasets/polity3/politymay96.data. Both data sets are produced independently from the democratic peace research program, and the initial codings, from the 1980s, precede it. Reviews of the program include Steve Chan, "In Search of Democratic Peace: Problems and Promise," *Mershon International Studies Review* 41, no. 1 (1997); James Lee Ray, "Does Democracy Cause Peace?" *Annual Review of Political Science* 1 (1997); and Bruce Russett and Harvey Starr, "From Democratic Peace to Kantian Peace: Democracy and Conflict in the International System," in Manus Midlarsky, ed., *Handbook of War Studies*, 2d ed. (Ann Arbor: University of Michigan Press, forthcoming).

[3] Henry Farber and Joanne Gowa, "Common Interests or Common Polities?" *Journal of Politics* 57, no. 2 (1997); Gowa, *Ballots and Bullets: The Elusive Democratic Peace* (Princeton: Princeton University Press, 1999); Douglas Lemke and William Reed, "Regime Types and Status Quo Evaluations," *International Interactions* 22, no. 2 (1996); Erik Gartzke, "Kant We All Just Get Along? Opportunity, Willingness and the Origins of the Democratic Peace," *American Journal of Political Science* 42, no. 1 (1998).

[4] The MIDs data (fn. 2) are unavailable after 1992, and data on dyadic trade are sparse and unreliable before 1885. In any event the further back one goes into the nineteenth century, the rarer are instances of democracy, intergovernmental organizations, and high levels of economic interdependence. The MIDs data include only disputes between recognized states and not, for example, extrasystemic (i.e., colonial) actions, covert operations, or domestic military interventions in support of a recognized government.

[5] We will not here offer a new theory on why democracy produces peaceful relations. A recent statement is Bruce Bueno de Mesquita et al., "An Institutional Explanation of the Democratic Peace," *American Political Science Review* 93, no. 4 (1999).

In keeping with the Kantian perspective, we expand our analysis beyond the democratic peace, incorporating the influence of economically important trade and joint memberships in international organizations. The classical liberals of the eighteenth and nineteenth centuries expected interdependence as well as popular control of government to have important pacific benefits. Commercial relations draw states into a web of mutual self-interest that constrains them from using force against one another. Thus interdependence and democracy contribute to what we have called the "liberal peace." Kant emphasized, in addition, the benefits of international law and organization. Our previous analyses of the cold war era indicate that, during those years at least, trade and networks of intergovernmental organizations did reduce the number of militarized interstate disputes; these effects were on top of the benefits of democracy.[6] We show here that they also operated in earlier and later years.

OUR OBJECTIVES AND METHOD

Although the liberal and realist perspectives are often considered antithetical, in keeping with Kant's philosophical analysis we conduct our tests of the Kantian peace while taking into account important realist influences. We believe, as Kant did, that both perspectives matter, as

[6] John R. Oneal and Bruce Russett, "The Classical Liberals Were Right: Democracy, Interdependence, and Conflict, 1950–1985," *International Studies Quarterly* 40, no. 2 (1997); Russett, Oneal, and David R. Davis, "The Third Leg of the Kantian Tripod: International Organizations and Militarized Disputes, 1950–85," *International Organization* 52, no. 3 (1998); Oneal and Russett, "Assessing the Liberal Peace with Alternative Specifications: Trade Still Reduces Conflict," *Journal of Peace Research* 36, no. 4 (1999). Here we extend this line of research in three ways: (1) providing a conceptual synthesis of Kantian and realist theories that treats conflict as inherent but subject to important constraints; (2) extending the temporal domain for trade and IGOs into the nineteenth century; and (3) assessing realist theories regarding the role of the hegemon and Kantian theories about systemic influences in a way that addresses, among others, constructivist and evolutionary perspectives on the international system. Note that the Kantian influences may be mutually reinforcing in a dynamic system of feedback loops, as suggested by Wade Huntley, "Kant's Third Image: Systemic Sources of the Liberal Peace," *International Studies Quarterly* 40, no. 4 (1996); and Russett, "A Neo-Kantian Perspective: Democracy, Interdependence, and International Organizations in Building Security Communities," in Emanuel Adler and Michael Barnett, eds., *Security Communities in Comparative Perspective* (New York: Cambridge University Press, 1998).

We and others have begun to address some of these links, such as greater trade between democracies, the possibility that trade is diminished between conflicting states, the effect of democracy, trade, and peace in increasing membership in international organizations, and the effect of conflict on democracy. On the first, see Harry Bliss and Russett, "Democratic Trading Partners: The Liberal Connection," *Journal of Politics* 58, no. 4 (1998), and James Morrow, Randolph Siverson, and Tessa Tabares, "The Political Determinants of International Trade: The Major Powers, 1907–90," *American Political Science Review* 92, no. 3 (1998); on the second, see Soo Yeon Kim, "Ties That Bind: The Role of Trade in International Conflict Processes" (Ph.D. diss., Yale University, 1998); on the third, see Russett, Oneal, and Davis (this fn.); and on the last, see Oneal and Russett, "Why 'An Identified Systemic Model of the Democratic Peace Nexus' Does Not Persuade," *Defence and Peace Economics* 11, no. 2 (2000).

both consider conflict and the threat of violence to be inherent in a anarchic world of sovereign states. The Hobbesian element of this ur derstanding is central to realist theory, but it is also deeply embedde in the liberal tradition. Kant accepted Hobbes's description of a state (war among nations and believed that a balance of power could prever war; but history has shown all too clearly, as most realists acknowledg that this "peace" is tenuous. Kant, however, was convinced that a ger uine, positive peace could be developed within a "federation" of liber: republics that rested more on the three Kantian supports—democrac interdependence, and international law and organizations—than o power politics. The pacific federation envisioned by Kant is not a worl state but a federation whose members remain sovereign, linked only b confederational or collective security arrangements. Liberalism, that i: sees democratic governance, economic interdependence, and interna tional law as the means by which to supersede the security dilemm rooted in the anarchy of the international system. For states not muc linked by these ties, however, the threat of violence remains. In addi tion, liberal states must fear those illiberal states that remain outside th Kantian confederation.[7]

Thus we begin by assuming that the international system is anarchi and power is important. Yet despite the inherent possibility of violence states do not fight all others or at all times even where realist principle dominate. Rather, they are constrained by power, alliances, and dis tance. States must be concerned with the balance of power and the co incidence of national interests expressed in alliances. Many states moreover, are irrelevant to these calculations: in general, the farthe apart two states are, the fewer are the issues over which to fight and th less the threat they pose to one another. Ultimately therefore realists ar concerned only with states that have the opportunity and incentive t engage in conflict.[8] Accordingly because these constraints provide baseline against which to assess the additional impact of the Kantiar influences, we incorporate them as central features of our theoretica model. To the realist variables we add measures for the three Kantiar constraints, hypothesizing that (1) democracies will use force less fre quently, especially against other democracies; (2) economically impor tant trade creates incentives for the maintenance of peaceful relations

[7] Michael W. Doyle, *Ways of War and Peace* (New York: W. W. Norton, 1997), chap. 8; David Lake "Powerful Pacifists: Democratic States and War," *American Political Science Review* 86, no. 4 (1992).

[8] Birger Heldt, "Inherency, Contingency, and Theories of Conflict and Peace" (Manuscript, Yal University, 1998); Benjamin Most and Harvey Starr, *Inquiry, Logic, and International Politics* (Colum bia: University of South Carolina Press, 1989), chap. 2.

and (3) international organizations constrain decision makers by promoting peace in a variety of ways. Since the modern international system is far from a pacific federation of democratic states, we expect both realist and Kantian factors to affect interstate relations. We explicitly consider how realist and liberal influences at both the dyadic and the systemic level have altered the functioning of the international system, addressing the role of the leading state and the influence of the changing Kantian variables over time.

Evidence for the pacific benefits of economic interdependence and membership in intergovernmental organizations (IGOs) is less widely accepted than is that for the democratic peace, and it has been subjected to less extensive critical scrutiny. We alone have assessed the effect of IGOs on conflict at the dyadic level of analysis. Moreover, theoretical expectations regarding the impact of trade and IGOs are more diverse than those concerning democracy. No one hypothesizes that democracies are *more* likely to fight each other than are other polities; but the liberal view of the pacific effects of trade is contradicted by those who expect conflict over the division of the gains from trade and by the dependency school and its intellectual predecessors and descendants, who expect conflict between large and small states.[9] As for IGOs, a plausible view might be that states form or join international organizations to manage—albeit often without success—disputes with their adversaries, the UN being an example. More commonly, realists regard international institutions as nearly irrelevant to the security issues at the heart of high politics, with no effect independent of existing power relations.[10] Even among those who hold that trade or IGOs play a positive role in promoting peace, the reasons advanced vary. Rational choice theorists emphasize political actors' complementary economic interests in maintaining peaceful interstate relations—interests that are reflected in the decisions of national leaders. Fearful of the domestic political consequences of losing the benefits of trade, policymakers avoid the use of force against states with which they engage in economically important trade. But one can also devise constructivist explanations about how the communication associated with trade builds cross-national sentiments of shared identity.[11]

[9] A useful review is Susan McMillan, "Interdependence and Conflict," *Mershon International Studies Review* 41, no. 1 (1997).

[10] John Mearsheimer, "The False Promise of International Institutions," *International Security* 19 (Winter 1994–95).

[11] Emanuel Adler and Michael Barnett, "Security Communities in Theoretical Perspective," in Adler and Barnett (fn. 6); Alexander Wendt, *Social Theory of International Politics* (New York: Cambridge University Press, 1999). For microlevel evidence that trading contacts expand elites' views of

Even realists acknowledge that international institutions like NATC help to preserve peace among their members by supplementing the deterrent effect of sheer military power. Liberals emphasize the potential of institutions for communicating information and facilitating bargaining,[12] while constructivists see institutions as instruments for expanding people's conceptions of identity, relatedness, and self-interest. Because IGOs vary widely in their functions and capabilities, any or all of these explanations may be correct in a particular instance.[13] As with the consequences of democratic institutions, we do not attempt to resolve these theoretical debates here. Instead we seek to offer an empirical assessment of the effect of the Kantian influences on interstate relations.

In expanding the historical domain of the Kantian peace, one encounters hurdles (and opportunities) that arise from marked changes in the nature of political regimes, the importance of international trade, and the role of international organizations. As measured by the standard data on political regimes, Polity III,[14] the average level of democracy in the international system has risen since the early 1800s, in a pattern that is sporadic and wavelike.[15] Similarly, the mean level of economic interdependence as measured by the ratio of bilateral trade to gross domestic product fell after World War I but rose again in subsequent years. Most clear is the growth in the number of IGOs, though those associated with the creation and sustenance of a truly global economy largely emerge only after World War II. These trends for the 1885–1992 period are shown in Figure 1.[16] Higher levels of democracy, interdependence, and IGO membership should, of course, reduce conflict for the pairs of countries affected; but we also expect that as the number of democracies increases, trade grows, and IGOs proliferate, there will be important systemic influences on other pairs of states as well. The effect of the Kantian influences should, we hypothesize, be apparent over time as well as cross-nationally.

---

their self-interest, see Daniel Lerner, "French Business Leaders Look at EDC," *Public Opinion Quarterly* 24, no. 1 (1956); and Bruce Russett, *Community and Contention: Britain and America in the Twentieth Century* (Cambridge: MIT Press, 1963), chap. 9.

[12] Robert O. Keohane and Lisa Martin, "The Promise of Institutionalist Theory," *International Security* 20, no. 1 (1995); Lisa Martin and Beth Simmons, "Theories and Empirical Studies of International Institutions," *International Organization* 52, no. 4 (1998).

[13] For a review of some relevant hypotheses and findings, see Russett, Oneal, and Davis (fn. 6).

[14] Jaggers and Gurr (fn. 2).

[15] Samuel P. Huntington, *The Third Wave: Democratization in the Late Twentieth Century* (Norman: University of Oklahoma Press, 1991).

[16] For graphing purposes the scale for bilateral trade/GDP has been increased by two orders of magnitude and that for IGO membership has been reduced by one order of magnitude.
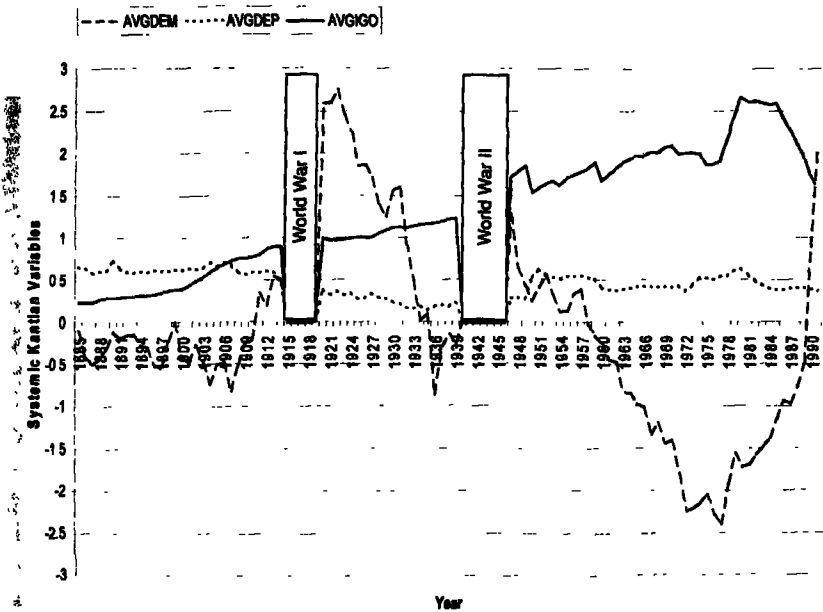
FIGURE 1
KANTIAN VARIABLES

Our statistical method—pooled cross-sectional time-series regression analysis of data regarding pairs of states (dyads) observed annually—is well suited to the purposes at hand. It considers variance in states' involvement in militarized disputes across dyads in each year and in dyadic relations through time. Consequently we can determine the likelihood of conflict as a function of differences across thousands of pairs of states annually and of changes in dyadic relations or in the international system from year to year over a period of more than one hundred years. By measuring change in the Kantian variables through time, we can begin to disentangle their systemic effects from their strictly dyadic influences.

Changes over time in the average level of democracy, interdependence, and IGO involvement capture important elements of the norms and institutions of the international system. Wendt, for instance, contends that world politics has slowly evolved from Hobbesian anarchy to a Lockean system wherein the security dilemma is ameliorated by norms recognizing the right of sovereign states to exist; these, in turn

the relative power of the United States (in its capabilities if not in its will to control or shape events). These theoretically based distinctions require us to consider the effects of the Kantian variables within different international structures, though evaluation of the post–cold war era necessarily remains tentative.

We omit from our analyses all but the first year of both World War I and World War II, because bilateral trade data for those years are fragmentary, as they are for the immediate postwar years, 1919–20 and 1946–49. Omitting all but the first year of the world wars, which consisted of conflicts between democracies and autocracies or between two autocracies, biases our results against finding evidence of the democratic peace, but it also provides assurance that our results are not determined by these dramatic but atypical events.[23] Most of our variables and data are discussed in previous publications. Here we concentrate on what is new.

DEPENDENT VARIABLE: INVOLVEMENT IN MILITARIZED DISPUTES

We use the Correlates of War (COW) data on interstate disputes (MIDs). We code each year that a dyad was involved in a dispute in which one or both states threatened to use force, made a demonstration of force, or actually used military force against the other. The variable DISPUTE equals 1 if a dispute was ongoing and 0 if not. Some researchers urge that only the initial year of a dispute be noted since a dispute in one year increases the chances that the dyad will experience a dispute in subsequent years.[24] This procedure eases some problems but raises others. If leaders are rational, as all our theories assume, they will frequently reevaluate their positions, whether to escalate, deescalate, or maintain the existing strategy. We agree with Blainey: "The beginning of wars, the prolonging of wars and the prolonging or shortening of pe-

---

effects of nuclear weapons would also imply a break between 1945 and all previous years of modern history. Dating the end of the bipolar cold war system is more problematic. Waltz's definition would argue for a break at the end of 1991, when the Soviet Union was dissolved. But William Dixon and Stephen Gaarder show a decisive shift in the pattern of Soviet-American conflict in 1988; see Dixon and Gaarder, "Presidential Succession and the Cold War: An Analysis of Soviet-American Relations, 1948–1992," *Journal of Politics* 54, no. 1 (1992).

[23] Farber and Gowa (fn. 3) express this concern.

[24] Stuart A. Bremer, "Dangerous Dyads: Conditions Affecting the Likelihood of Interstate War," *Journal of Conflict Resolution* 36, no. 1 (1993); Katherine Barbieri, "International Trade and Conflict: The Debatable Relationship" (Paper presented at the annual meeting of the International Studies Association, Minneapolis, Minn., February 1998); Nathaniel Beck, Jonathan Katz, and Richard Tucker, "Taking Time Seriously in Binary Time-Series-Cross-Section Analysis," *American Journal of Political Science* 42, no. 4 (1998). See, however, our comment in fn. 49 below.

riods of peace all share the same causal framework. The same explanatory framework and the same factors are vital in understanding each state in the sequel of war and peace."[25] Moreover, we investigated 166 multiyear disputes during the post–World War II era and found that more than half involved a change in the level of force employed over the course of the dispute or that a new dispute arose as the first was concluding. Thus we report analyses of states' involvement in disputes rather than of just their onset; but as in earlier studies of the cold war era,[26] we reestimated key analyses using only the first year of disputes without finding material differences from those reported below.

Dyadic Independent Variables

We lag all independent variables by one year to ensure that they were not affected by a dispute to be explained. For some explanatory variables this precaution is clearly important; for example, conflict may limit trade just as trade may constrain conflict. A similar reciprocal relationship can be imagined for international organizations and conflict, as many IGOs—though hardly all—are formed among states that maintain peaceful relations. For other variables such considerations are irrelevant. Geographically proximate countries are prone to conflict, but the frequency of their disputes does not affect their proximity. To be consistent, however, we lag all the independent variables. This precaution does not put to rest all questions about the direction of causality, but it is a reasonable step at this time.[27] All the variables are listed by their acronyms in the appendix.

DEMOCRACY

We use the Polity III data to compute a summary measure of the political character of regimes, subtracting every country's score on the autocracy scale from its score on the democracy scale. The resulting variable (DEM_i) ranges from −10 for an extreme autocracy to +10 for the most democratic states. Because a dispute can result from the actions of a single state, the likelihood of conflict is primarily a function of the degree of constraint experienced by the less constrained state in each dyad. As that state is the weak link in the chain of peaceful

[25] Geoffrey Blainey, *The Causes of War*, 3d ed. (New York: Free Press, 1988).

[26] Oneal and Russett (fn. 6, 1999).

[27] Kim (fn. 6), using a simultaneous equation model, finds that the effect of trade on conflict is much stronger than the reciprocal one. Russett, Oneal, and Davis (fn. 6) construct a model for predicting IGO membership that includes, among other factors, the absence of conflict. There is an effect, but it is weaker than the influence of IGOs on conflict.

relations,[28] we expect that the less democratic state (DEM_L) in a dyad is the stronger determinant of interstate violence. Conversely, the more democratic that state, the more constrained it will be from engaging in a dispute and the more peaceful the dyad. In previous analyses we found, as Kant had expected, that the *difference* between states' political regimes also affects the likelihood of conflict. Democratic-autocratic dyads were the most conflict-prone in the cold war era; two autocracies were less likely to fight, and two democracies were the most peaceful. We reconsider these findings below.

The Polity III regime scores exhibit some problems of comparability over time. Until 1918 about 40 percent of British males (disproportionately working class) were disfranchised by residence requirements; female suffrage was granted partially in 1918 and fully only in 1928.[29] In the United States women obtained the vote only in 1920, and blacks were systematically excluded until the 1960s. Swiss women achieved the franchise only in 1971. Some of these changes are reflected in the Polity data and hence in rising levels of democracy in the international system. For example, the United Kingdom goes from 6 to 7 on the democracy scale in 1880, to 8 in 1902, and jumps to 10 only in 1922. But Switzerland is coded at 10 from 1848, as is the United States from 1871. The consequences of these restrictions on political participation for foreign policy may not be trivial. In the contemporary United States, for example, women are significantly more averse to the use of military force than are men and vote in part on this basis.[30] Thus the exclusion of women from the franchise in earlier periods could have profoundly reduced the tendency of even the most "democratic" states to avoid conflict.

[28] William J. Dixon, "Democracy and the Peaceful Settlement of International Conflict," *American Political Science Review* 88, no. 1 (1994).

[29] Trevor Wilson, *The Myriad Faces of War: Britain and the Great War, 1914–1918* (Cambridge, England: Polity Press, 1986), 660–61; Kenneth MacKenzie, *The English Parliament* (Harmondsworth: Penguin, 1950), 106.

[30] Carole Kennedy Chaney, R. Michael Alvarez, and Jonathan Nagler, "Explaining the Gender Gap in U.S. Presidential Elections," *Political Research Quarterly* 51, no. 2 (1998). To take such changes into account, Zeev Maoz uses an adjusted threshold of democracy for all countries that shifts upward in 1870 (for general male suffrage) and 1920 (female suffrage); see Maoz, *Domestic Sources of Global Change* (Ann Arbor: University of Michigan Press, 1996), 54. Our use of unadjusted democracy scores thus leans against our hypothesis of democratic peace before World War I. Kristian Gleditsch and Michael Ward note that our continuous measure, Democracy minus Autocracy score, has the virtues of being symmetric and transitive; but the relative importance of its components is unstable over time; see Gleditsch and Ward, "Double Take: A Re-examination of Democracy and Autocracy in Modern Polities," *Journal of Conflict Resolution* 41, no. 3 (1997). For the period 1880–1969 this aggregated measure is largely influenced by the degree of competition for executive recruitment; subsequently constraints on the executive are the main determinant. Fortunately the relatively stable earlier period covers all the pre–cold war years we add here. As no analysis of the democratic peace after World War II has yet addressed the 1969 break, we too leave that for later investigation.

ECONOMIC INTERDEPENDENCE

For most of the post–World War II era the measurement of this Kantian variable is straightforward, because the International Monetary Fund reports statistics regarding bilateral trade. Since trade is expected to influence dyadic relations only if it is economically important, we divide the sum of a country's exports and imports with its partner by its GDP, as reported in the standard references for the years after 1950.[31] As with the influence of democratic institutions, we expect the likelihood of a dispute to be primarily a function of the freedom of the less constrained state to use force, that is, the bilateral trade-to-GDP measure of the state less dependent economically on trade with its dyadic partner (DEPEND_L). We also report tests for a positive effect of asymmetric dependence on conflict, as proposed by dependency theorists.

When we move back to the years before World War II, however, national economic data become more problematic. During the years 1920–38 the League of Nations compiled contemporary data on bilateral trade in current values, along with exchange rates.[32] While the accuracy and the comparability of these data are undoubtedly less than in the later IMF reports, they are the best available. There are no institutional compilations of trade data for the years of the two world wars, nor for the period before 1914. Before World War I the annual editions of *The Statesman's Yearbook*[33] offer the closest approximations, but these data are less standardized, the appropriate exchange rates for converting the data to a common unit are less certain, and more data are missing.[34]

[31] International Monetary Fund, *Direction of Trade (ICPSR 7623)* (Washington, D.C.: IMF, 1993; distributed by Ann Arbor, Mich.: Inter-University Consortium for Political and Social Research). Robert Summers et al., *The Penn World Table (Mark 5.6a)* (Cambridge, Mass.: National Bureau of Economic Research, 1995). Due to missing data for trade and/or GDP, the great majority of dyads involved in the Korean and Vietnam Wars are omitted, as are most Arab-Israeli dyads. Since most of those are conflicting democratic-autocratic dyads with no trade, our analysis is likely to be biased against the liberal hypotheses. Because these conflicts spanned several years, excluding these cases mitigates the problem of temporal dependence in the time series, as does omitting all but the first year of the world wars. Also omitted are roughly 2,500 communist dyad-years: non-IMF members. These states traded among themselves but did not report it to the IMF and generally had little conflict. Had we been able to include them, the post–1950 sample would have been increased by only about 2 percent.

[32] League of Nations, *International Trade Statistics* (Geneva: League of Nations, annual volumes).

[33] Martin Epstein, ed., *The Statesman's Yearbook, 1913* (London: Macmillan, 1913), and earlier annual editions by other editors.

[34] We took several steps to minimize missing trade data in this period. We used information regarding one state's exports to another to infer its partner's imports; we interpolated between known values of trade and used the average value of a dyad's trade to extrapolate; and we assumed, for those states for which we had data, that there was no trade between any two if neither reported any exports or imports with the other. As a result we have trade data for 61 percent of the dyads 1885–1913 and 1920–38. We conducted several tests to see if these methods might have biased our results. First we dropped all zero values of trade, and then we dropped all interpolations and extrapolations. Analyses with the remaining "real" data, 1885–1940, revealed little change in the results. We also determined

Because of these difficulties we collected alternative estimates for bilat eral trade in the 1885–1949 period, compared them with the data fror *The Statesman's Yearbook* and the League of Nations, and adjusted th data from our principal sources as appropriate.[35]

Information on dyadic trade, the numerator of the dependence mea sure, is only half of the problem for the pre-1950 era, however. To cal culate the economic importance of trade we need estimates of nation gross domestic products. No comprehensive collection of GDP data ex ists, but Maddison provides estimates in constant dollars for fifty-si countries in all regions of the world for 1870–1992.[36] We used these i a two-step procedure to estimate the GDPs in current dollars for a larg number of countries. First, we regressed Maddison's constant dolla GDP estimates on states' total annual energy consumption, the regio where they were located, the year, and various interactive terms. Annua energy consumption, collected by the Correlates of War (COW) project is a good correlate of the size of national economies, as Morgenstern Knorr, and Heiss noted twenty-five years ago.[37] More than 93 percen of the variance in Maddison's GDPs was explained. Based on the coeffi cients in this analysis, we were able to estimate the constant dollar GDP for a large number of other countries. Second, we converted these con stant dollar estimates to current dollars, using Maddison's U.S. dolla GDP deflator.

JOINT IGO MEMBERSHIPS

The influence of international organizations on interstate conflict, th last Kantian variable, is assessed by the number of IGOs in which bot

---

that the sample of dyads for which we have trade data is unlikely to be biased. To do this, we created variable (MISSING) that equaled 1 if DEPEND_L was missing and 0 otherwise and then changed all miss ing values of DEPEND_L to zero. We then estimated equation 1 below with the variable MISSING added It was not statistically significant, indicating that the incidence of disputes among the dyads for whic trade (or GDP) data are missing does not differ from that for the dyads for which data are availabl

[35] These include volumes by Brian R. Mitchell for each region of the world and for the Unite Kingdom (Cambridge: Cambridge University Press, various years); U.S. Department of Commerc *Historical Statistics of the United States: Colonial Times to 1970* (New York: Basic Books, 1976); an Katherine Barbieri's data posted at http://pss.la.psu.edu/TRD_DATA.htm. Exchange rates come fro U.S. Federal Reserve Bank sources, *The Statesman's Yearbook*, and Global Financial Data Compan www:globalfindata.com.

[36] Angus Maddison, *Monitoring the World Economy, 1820–1992* (Paris: Organization for Economi Cooperation and Development, 1995). His U.S. dollar GDP deflator is found in Maddison, "A Lon Run Perspective on Saving" (Manuscript, Institute of Economic Research, University of Groningen October 1991).

[37] Oskar Morgenstern, Klaus Knorr, and Klaus P. Heiss, *Long Term Projections of Power: Politica Economic, and Military Forecasting* (Cambridge, Mass.: Ballinger, 1973); and also John R. Oneal, "Mea suring the Material Base of the East-West Balance of Power, *International Interactions* 15, no. 2 (1989)

:s in a dyad share membership, as reported by the *Yearbook of Inter-* *onal Organizations.*[38] Simply counting joint memberships (ranging ı 0 to over 130 for some dyads in recent years) is far from an ideal sure of the importance and effectiveness of international organiza-s. It includes organizations that are weak and strong, regional and al, functional and multipurpose. Ideally the total should be broken n and some organizations given special weight, but this is hard to s a practical matter and there is little theory to guide the attempt. now we use the simple count of joint memberships in intergovern-tal organizations; this variable is labeled IGO.

\BILITY RATIO

first of the realist constraints on states' use of military force is rel-: power, specifically the balance of power within a dyad. The idea an equal balance of power may deter conflict has deep historical s, as does the idea that a preponderance of capabilities is more y to preserve the peace by reducing uncertainty as to which side ld win a contest of arms. Recent empirical work suggests, however, it is preponderance that deters military action.[39] Our index of rel-: power (CAPRATIO) is the natural logarithm of the ratio of the ıger state's military capability index to that of the weaker member ıch dyad. We make these calculations using the COW data on pop-on, industry, and military forces.[40]

ANCE

:s are generally thought to fight each other less than other states use they share common security interests. They often share other :ical and economic interests as well. We control for this influence g a variable (ALLIES) that equals 1 if the members of a dyad were :d by a mutual defense treaty, neutrality pact, or entente; it equals 0 rwise.[41]

Ve extended the data from the sources in Russett, Oneal, and Davis (fn. 6).
ıremer (fn. 24); Jacek Kugler and Douglas Lemke, eds., *Parity and War: Evaluations and Exten-* *f the War Ledger* (Ann Arbor: University of Michigan Press, 1996). Waltz (fn. 22), 117–23, re-the balance of power literature and states his own version.
›ata are from J. David Singer and Melvin Small, *National Military Capabilities Data* (Ann Arbor: rsity of Michigan, Correlates of War Project, 1995); the date of final modification of the data ecember 28, 1994.
Ve updated J. David Singer, *Alliances, 1816–1984* (Ann Arbor: University of Michigan, Corre-f War Project, 1995), with material from N. J. Rengger, with John Campbell, *Treaties and Al-* *of the World*, 6th ed. (New York: Stockton, 1995).

CONTIGUITY AND DISTANCE

The potential for interstate violence exists when at least one member o
a dyad can reach the other with effective military force. For most state
the ability to do so is determined foremost by geographic proximity, es
pecially as one goes farther back in history. Furthermore, neighbors ar
likely to have the most reasons to fight—over territorial boundaries
natural resources, the grievances of cross-border ethnic groups, and s
on. Thus the constraint of distance reduces the capability to fight an
most of the incentives to do so as well; this finding is extremely stron
in previous research.

Accordingly, we include two different terms in our regression analy
ses to capture this effect as fully as possible. DISTANCE is the natura
logarithm of the great circle distance in miles between the capitals o
the two states (or between the major ports for the largest countries)
using the logarithm acknowledges a declining marginal effect. Addi
tionally we include NONCONTIG, a measure that equals 1 if two state
are not directly or indirectly contiguous (via colonies or other depen
dencies). It equals 0 if they share a land boundary or are separated b
less than 150 miles of water. Because of the widespread nature of colo
nial empires, these two measures are not highly correlated (r^2 = 0.21)
especially up to World War II. The effect of distance in constrainin
conflict, however, is less for the great powers: those with the land, sea
or (in the last half-century) air capability to deliver substantial forces o
destructive power globally. The COW project has identified these majo
powers on the basis of a consensus of historians. To give full consider
ation to realists' concerns, we add a third variable, MINORPWRS, coded 1
if a dyad is composed of minor powers and 0 for those that include a
least one great power. (To be consistent with our view that conflict i
endemic but subject to constraints, we reverse the terminology an
coding of the last two variables from those in previous research report
where we used CONTIG and MAJOR. This has no effect on our statistica
analyses, other than to reverse the sign of the coefficients. Note tha
some contiguous dyads also include one or two major powers.)

In most of the analyses below, we include all possible pairs of state
for which information is available, using COW data regarding member
ship in the international system to generate these cases. Thus we do no
limit our study to the politically relevant dyads, identified as contigu
ous states and dyads containing at least one major power. We continu
to believe that such a restriction makes good theoretical sense, however
These dyads are much more likely to engage in military disputes. Polit-

ically relevant dyads constitute just 22 percent of all the dyads for which we have data; nevertheless they account for 87 percent of all the disputes. In other words, the politically relevant dyads are twenty-four times more likely to experience a militarized dispute than are those we have deemed to be "irrelevant." And some disputes among these other dyads are contagion effects of being drawn into conflicts through alliance commitments. We include all dyads in most of the analyses reported below to be sure we are not ignoring the causes of these other disputes,[42] but we also explore the consequences of including the nonrelevant pairs.

Systemic Independent Variables

KANTIAN SYSTEMIC AND RELATIVE DYADIC MEASURES

To clarify the influence of the international system on the likelihood of dyadic conflict, we create three system-level Kantian variables and three realist variables, the latter designed to capture the hegemon's effect on interstate relations. The three Kantian variables are straightforward derivations of our basic measures: we simply computed the means of DEM, DEPEND, and IGO for each year. These are the measures (omitting the years of the world wars) graphed in Figure 1. In the analyses below, they are identified as AVGDEM, AVGDEPEND, and AVGIGO. We hypothesize that the greater these systemic measures, the more the global system will reflect the normative and institutional constraints associated with democracy, interdependence, and the rule of law. It is also possible to assess the standing of each dyad in each year relative to our three annual Kantian averages. Thus we calculated three relative dyadic measures: $RELDEM_L$ = (DEM_L – AVGDEM) / the standard deviation of DEM; $RELDEPEND_L$ = ($DEPEND_L$ – AVGDEPEND) / the standard deviation of DEPEND; and RELIGO = (IGO – AVGIGO) / the standard deviation of IGO. These measures identify the dyads that were most democratic, interdependent, and involved in intergovernmental organizations at each point in time. By dividing by the standard deviations, we can directly compare these estimated coefficients. Combining systemic and relative measures in a single equation allows us to compare the effect of changing values of the Kantian variables through time versus the standing of dyads cross-sectionally relative to

[42] As recommended by William Reed, "The Relevance of Politically Relevant Dyads"(Paper presented at the annual meeting of the Peace Science Society [International], New Brunswick, N.J., October 1998).

the annual means. We expect the systemic and relative variables to make independent contributions to the frequency of dyadic disputes.

REALIST SYSTEMIC MEASURES

Hegemony. We also create three systemic variables associated with prominent realist theories regarding the hegemon's influence on international relations. Hegemonic-stability theory postulates that the most powerful state in the system, the hegemon, has the ability to constrain weaker states from resorting to violence.[43] This power to keep the peace might be manifested as dominance within the hegemon's sphere of influence and the ability to deter adversaries from using military force in a way detrimental to its interests. A crude but reasonable measure of the power of the leading state is its share of all the major powers' capabilities in each year. As before, we use COW data to make this calculation.

Identification of the hegemon is not obvious in all cases. Through much of recent history it is not clear whether any state was truly hegemonic.[44] It is generally agreed that in the thirty years before World War I the United Kingdom was closer than any other country to being hegemonic, although its power relative to both Germany and the United States was declining. During the interwar era the United States clearly had greater economic strength and military potential than the United Kingdom; but its actual military power was only about equal. Moreover, its geographic position and isolationist policy limited its involvement in the Central European system. Consequently, we accept Organski and Kugler's judgment that Britain was the hegemon in the interwar period as well.[45] In the post–World War II years, if any state can be said to have been hegemonic, it is the United States. Hence we use the proportion of capabilities held by the United Kingdom as the measure of the hegemon's power in the first sixty years analyzed and that of the U.S. after 1945. Our systemic indicator (HEGPOWER) has reasonable face validity, declining from 33 percent in 1885 to 14 percent in 1913, and dropping under 11 percent by 1938. America's hegemony is manifest immediately following World War II, when it controlled 52 percent of the major powers' capabilities. This declined

[43] Robert Gilpin, *War and Change in World Politics* (Cambridge: Cambridge University Press, 1981).

[44] Bruce Russett, "The Mysterious Decline of American Hegemony, or, Is Mark Twain Really Dead?" *International Organization* 32, no. 2 (1985).

[45] A. F. K. Organski and Jacek Kugler, *The War Ledger* (Chicago: University of Chicago Press, 1980) On measurement, see David Sacko, "Measures of Hegemony" (Paper presented at the annual meeting of the Peace Science Society [International], New Brunswick, N.J., October 1998).

26 percent by the early 1980s but rose to 29 percent with the collapse the Soviet Union.

Satisfaction with the status quo. The power-transition theory origi-lly advanced by Organski consists of propositions not only about the nstraining influence of an imbalance of power but also about the role ayed by states' satisfaction with the status quo. States rising in power ill challenge a hegemon only if they are dissatisfied with the interna-onal system it dominates. Lemke and Reed extend this rationale in an fort to subsume the democratic peace within power-transition the-ry.[46] They contend that democracies have fought less historically cause the hegemon has been democratic since the end of the apoleonic Wars. First Britain and then the United States, it is argued, ed its power to construct an international system that provided nefits to itself and its mostly democratic allies. Thus democracies' tisfaction with the status quo created by the most powerful demo-atic state and reinforced by its system of alliances accounts for the ace among democratic dyads. Like Lemke and Reed, we assess this w by computing a measure of each state's satisfaction with the status o based on the correspondence between its portfolio of alliances and at of the hegemon, as indicated by the tau-b measure of statistical as-ciation. Then we multiply the scores of the two states in a dyad to ate a measure of joint satisfaction (SATISFIED).[47] This measure indi-es the degree to which each dyad is content with the distribution of efits achieved under the leadership of the dominant state.

Hegemonic tensions. Both hegemonic-stability theory and power-nsition theory hold that the international system will be more peace-when the hegemon is strong relative to its principal rivals. The gemon may also affect the system by transmitting concerns for its n security to other states. International tensions involving the hege-nic power are likely to have consequences for its allies, its rivals, its als' allies, and even neutral states. "When elephants fight, the grass ts trampled," as the adage goes. It is also possible, to extend the taphor, that when small animals fight, big ones will be drawn in. rge states may intervene in ongoing conflicts because they see an op-rtunity to achieve gains or avoid losses. Either way, international ten-

Lemke and Reed (fn. 3).

We added 1 to each state's tau-b score to make it positive. The tau-b index of the similarity of al-ce portfolios was introduced by Bruce Bueno de Mesquita, "Measuring Systemic Polarity," *Journal onflict Resolution* 19, no. 2 (1975). It was adapted as a dyadic measure of satisfaction by Woosang "Alliance Transitions and Great Power War," *American Journal of Political Science* 35 (1991), and ently used by Lemke and Reed (fn. 3).

sions may be contagious. To assess this view, we created a measure o the hegemon's sense of its own security, calculating its defense spendin as a share of its GDP (HEGDEF).[48] We hypothesize that the global systen will experience more numerous disputes when the hegemon is commit ting more of its resources to the military. In such times, the hegemon pre sumably perceives greater threats to its interests. To assess the scope o contagion, we consider whether involvement in disputes rises mostly fo the hegemon itself, for the hegemon and its allies, or for unallied states.

RESULTS

We evaluate the Kantian peace, 1885–1992, employing logistic regres sion analysis. First we assess the effects of democracy, interdependence and IGOs using a simple dyadic specification. In this view the likelihoo of conflict is primarily determined by the state less constrained eco nomically or politically. We also consider the degree to which the polit ical and economic characteristics of the other member of a dyad affec the likelihood of a militarized dispute. Next we disentangle the sys temic and cross-sectional influences of the Kantian variables on dyadi conflict. We consider the effects of trends in the underlying variable and each dyad's degree of democracy, interdependence, and involve ment in IGOs relative to these annual systemic averages. Finally we in vestigate central realist tenets regarding the role of the leading state i the international system.

We examine the involvement in militarized interstate disputes c nearly 6,000 pairs of states observed annually, for a total of almo 150,000 observations. Because of the lagged variables the analysis be gins with disputes in the year 1886 that are explained by reference t conditions in 1885. As noted earlier, we do not consider the two worl wars after the first year of conflict or the immediate postwar years; th is, we exclude disputes for 1915–20 and 1940–46.

Unless otherwise indicated, we estimate the coefficients in our re gression equations using the general estimating equation (GE method. We adjust for first-order autoregression (AR1) and estima statistical significance using robust standard errors that take into a count the clustering of our data by dyads. Thus we respond to the co

[48] Military expenditure is a component of the COW index of militarily relevant capabilities. On t validity of our measure, see John R. Oneal and Hugh Carter Whatley, "The Effect of Alliance Me bership on National Defense Burdens, 1953–88," *International Interactions* 22, no. 2 (1996). Chan in this index for the hegemon's military burden correlate highly with changes in the average milit burden for all the major powers.

cerns raised by Beck, Katz, and Tucker. We rely on GEE rather than on their recommended solution for temporal dependence because of doubts about its appropriateness, especially given the strong, theoretically specified relation between trade and the time elapsed since a dyad's last dispute.[49] We have, however, reestimated our key equations using their method as a check on our findings. Because our hypotheses are directional and we have corrected for these violations in the assumptions underlying regression analysis, we report one-tailed tests of statistical significance.

EVALUATING THE KANTIAN PEACE USING THE WEAK-LINK SPECIFICATION

Our first test is the simplest. We expect the likelihood of conflict to be primarily a function of the degree to which the less constrained state along each of several dimensions is free to use military force. This is the weak-link assumption that this state is more likely to precipitate a break in the peace: the less the political or economic constraints on that state's use of force, the greater the likelihood of violence. Consequently we include the lower democracy score and the lower bilateral trade-to-GDP ratio. The number of joint memberships in international organizations is inherently a dyadic measure; it completes the Kantian specification. We include in the regression equation a measure of the dyadic balance of power and an indicator of whether the members of a dyad are allied. We also control for the distance separating the two states, whether or not they are contiguous, and whether both are minor powers.[50] Our first equation then takes the form:

$$\text{DISPUTE} = \text{DEM}_L + \text{DEPEND}_L + \text{IGO} + \text{ALLIES} + \text{CAPRATIO} + \text{NONCONTIG} + \text{DISTANCE} + \text{MINORPWRS} \quad (1)$$

The results of estimating equation 1, found in the first column of Table 1, provide strong support for the pacifying influence of democ-

[49] On GEE, see Peter J. Diggle, Kung-Yee Liang, and Scott L. Zeger, *Analysis of Longitudinal Data* (Oxford: Clarendon Press, 1994). We used the computing algorithms in StataCorp, *Stata Statistical Software*, Release 5.0 (College Station, Tex.: Stata Corporation, 1997). For Beck, Katz, and Tucker's methods, see fn. 24. We express our doubts that the effects of the theoretical variables and of time are separable, as Beck, Katz, and Tucker's method requires, in Oneal and Russett (fn. 6, 1999). GEE allows for temporal dependence in the time series but gives the theoretical variables primacy in accounting for interstate disputes. Beck, Katz, and Tucker introduce the PEACEYRS variables into the estimation process as coequals of the theoretical variables. See also D. Scott Bennett, "Parametric Methods, Duration Dependence, and Time-Varying Data Revisited," *American Journal of Political Science* 43, no. 1 (1999).

[50] Our recent specifications are found in Oneal and Russett (1997); and Russett, Oneal, and Davis (fn. 6). The controls, from Oneal and Russett (fn. 6, 1999), draw on Barbieri (fn. 24).

TABLE 1

MODELS OF THE KANTIAN PEACE, 1886–1992: PREDICTING INVOLVEMENT IN MILITARIZED DISPUTES

Variable	*1. 1886–1992 Simplest, All Dyads*	*2. 1886–1992 Peaceyears Correction*	*3. 1886–1939 All Dyads*	*4. 1886–1992, Politically Relevant Dyads*
Lower democracy	−0.0658***	−0.0628***	−0.0568***	−0.0595***
(DEM_L)	(0.0106)	(0.0093)	(0.0106)	(0.0106)
Trade/GDP	−57.8650***	−31.0726**	−43.2490**	−35.2394**
($DEPEND_L$)	(15.4901)	(10.6036)	(16.2861)	(12.3044)
International organizations	−0.0010	0.0160†	0.0068	−0.0068*
(IGO)	(0.0379)	(0.0042)	(0.0068)	(0.0039)
Capability ratio	−0.2337***	−0.1913***	−0.3638***	−0.2747***
(CAPRATIO)	(0.0502)	(0.0401)	(0.0664)	(0.0516)
Alliances	−0.2511	−0.3691**	−0.1727	−0.2822*
(ALLIANCES)	(0.1659)	(0.1574)	(0.1905)	(0.1677)
Noncontiguity	−2.0038***	−1.5864***	−1.3357***	−1.118***
(NONCONTIG)	(0.1836)	(0.1532)	(0.1844)	(0.1724)
Log distance	−0.4647***	−0.3615***	−0.3536***	−0.2610***
(DISTANCE)	(0.0571)	(0.0498)	(0.0620)	(0.0605)
Only minor powers	−1.8392***	−1.7208***	−1.8342***	−0.6754***
(MINORPWRS)	(0.1706)	(0.1351)	(0.1904)	(0.2082)
Constant	−1.9349***	−1.6174***	−2.2235***	−1.5765***
	(0.4731)	(0.4060)	(0.5316)	(0.4992)
Chi^2	1354.80	1920.45	494.98	193.43
P of Chi^2	0.0000	0.0000	0.0000	0.0000
Log Likelihood		−5732.4260		
Pseudo R^2		0.284		
N	149,373	149,404	33,346	33,334

*$p<.05$; **$p<.01$; ***$p<.001$, one-tailed tests; †$p<.001$, one-tailed test but wrong sign

y and trade: the more democratic the less democratic state in a dyad d the more economically important is trade, the greater is the likeliod of peace. The lower democracy and dependence measures are th significant at the .001 level. The number of joint memberships in Os, however, does not have a statistically significant effect on conflict this specification ($p < .40$). This is a consequence of two things: the lusion of all possible dyads in the analysis (not just those thought to politically relevant) and the rapid growth in the number of internanal organizations over time. The realist variables perform generally expected, though the indicator of alliance is only significant at the level: (1) a preponderance of power rather than a balance deters nflict; (2) contiguous states are prone to fight, as are those whose nelands are geographically proximate; and (3) major powers are inved in disputes more than are smaller states. All these variables are nificant at the .001 level.[51] Using the onset (or first year only) of a pute as the dependent variable produced nearly identical results.

Column 2 of Table 1 shows the results of estimating equation 1 ng Beck et al.'s correction for temporal dependence in the time se. The coefficients and significance levels are usually similar. The t notable exception involves the variable IGO. Its coefficient is now only positive but nearly four times its standard error.[52]

To see if the pacific benefits of the Kantian variables are limited to cold war era, we first reestimated equation 1 for just the early years, 6–1914 and 1921–39 using GEE. The results appear in column 3. paring them with column 1 shows much the same pattern as the ysis for all years. Both the lower democracy score ($p < .001$) and the ller bilateral trade-to-GDP ratio ($p < .004$) are highly significant.

o test the robustness of these results, we estimated separate regressions for each theoretically inng variable with just the controls for distance, contiguity, and major-power status. The signs and icance levels were consistent with those in the multivariate regressions, with one exception. Joint nemberships significantly ($p < .001$) reduced conflict in the restricted analysis. We also reestid equation 1 after dropping the measure of economic interdependence because this variable has ost missing values. The pacific benefits of democracy remained strong ($p < .001$). Joint membern IGOs, too, was significantly associated with a reduction in conflict ($p < .02$) when $DEPEND_L$ was ed. Not surprisingly, interdependent states share memberships in international organizations.

We suppress coefficients for the four spline segments to save space. All are significant ($p < .001$). s equation, and others presented subsequently, the coefficients for IGOs are the only ones not ro o the different methods for adjusting for temporal dependence. As our results suggest, joint ership in IGOs is most correlated of the three Kantian variables with the years of peace since a last dispute. Our methodological preference for GEE preceded our work on IGOs. We also estid equation 1 using conditional or fixed effects logistic regression. Greater democracy ($p < .001$) terdependence ($p < .05$) continued to be associated with peaceful dyadic relations, as was the exe of an alliance. Joint membership in IGOs and a greater capability ratio increased the prospects flict. These results are based on the 20,289 observations for dyads that experienced at least one e; 129,092 cases were dropped because the dependent variable always equaled zero.

Democracy and interdependence had strong peace-inducing effects during the multipolar period after 1885 and before the cold war. The benefits of democracy are strongest in the interwar years, but, as Gowa[53] also reports, by the decade leading to World War I democracies had become less likely to engage in militarized disputes with each other—an important shift that is obscured by using the years 1886–1914 as the period of analysis. In light of this evidence, the absence of democratic peace in the nineteenth century—not its presence in the cold war era—becomes the anomoly to be explained. The answer may lie more in the lower inclusiveness of democratic politics in that century than in characteristics of the international system.

Our measure of joint memberships in IGOs is insignificant for the period 1885–1939. The other coefficients in equation 1 are reasonably similar for the early years and the entire period. The effect of alliances before 1940 is even weaker ($p < .19$) than when all years are considered.

We also estimated equation 1 after creating an indicator for the 1989–92 post–cold war years and forming interactive terms with each of the three Kantian variables. The results indicate that the influence of democracy has not changed in this short span of time and the benefits of interdependence have been reduced, but IGOs are more important constraints on the threat or use of force.

In the past we limited our analyses to the politically relevant states, in the belief that the relations of most other dyads are not importantly influenced by the political and economic influences we have modeled. To see how including all possible pairs of states affects our results, we reestimate equation 1 using just the contiguous pairs of states and those that contain at least one major power—the politically relevant dyads—for all years, 1885–1992. This excludes dyads that in the great majority of cases had no reasonable opportunity to engage in armed conflict because the states were too far apart and had few issues over which to fight.

The last column of Table 1 provides strong support for the pacific benefits of all three elements of the Kantian peace. For the dyads most prone to conflict, joint membership in international organizations does reduce the likelihood of conflict ($p < .04$). The benefits of democracy ($p < .001$) and interdependence ($p < .002$) remain apparent. These results for the extended period, 1886–1992, are consistent with those in Russett, Oneal, and Davis, where only the years 1950–85 were considered; and they are more significant statistically.[54]

[53] Gowa (fn. 3), 98–100.

[54] Oneal, Russett, and Davis (fn. 6). Farber and Gowa (fn. 3), 409, analyze lower-level MIDs for 1816–1976 and find that democracy significantly affects the likelihood of conflict only after 1919.

Our tests with all possible pairs understate the pacific benefits of IGOs because most of these dyads do not have significant political-military relations. The probability that a nonrelevant dyad will become involved in a dispute is only 1/18 that of a major-power pair; it is 1/44 that of a contiguous dyad. Democracy, interdependence, and involvement in IGOs constrain states from using force; but if there is no realistic possibility of two states engaging in conflict, then the absence of these constraints will not increase the incidence of violence. With all dyads included a large number of false negatives obscures the hypothesized relationship. The theoretically interesting variables in equation 1 are simply irrelevant in explaining the state of relations, such as they are, between Burma and Ecuador, for example. Including numerous irrelevant dyads can bias the results, as we have recently shown with regard to trade.[55]

With logistic regression, the easiest way to show the substantive effects of the variables is to estimate the probability of a militarized dispute for various illustrative dyads. The same procedure is often used in epidemiological studies. For example, epidemiologists report the effect of various risk factors on the probability that an individual will contract lung cancer. As in our analyses, some of their independent variables are not subject to intervention (for example, age, heredity, gender; and for us distance and contiguity), while others are amenable to some degree of "policy" control (for example, diet, exercise, smoking; and for us alliances, democracy, interdependence, and IGOs). By statistical inference they, and we, can estimate the reduction in the probability of an event occurring if any one risk factor for a typical individual were different by a given amount.

For this, we calculated a baseline probability against which to make comparisons. We assumed the dyad is contiguous, because these states

However, using interactive terms for years, we find evidence of democratic peace by 1900. Earlier than that even the most democratic states were not democratic by contemporary standards. As democracy developed, the common interests of democracies and their antagonisms with authoritarian states may have become more substantial. Support for the benefits of democracy in Farber and Gowa's analyses is weakened by their decision to exclude consideration of all years of the world wars. Due to possible simultaneity problems, they do not control for alliances. Since alliances show little impact in our analyses, this may not matter. For results for trade that agree with ours, see Christopher Way, "Manchester Revisited: A Theoretical and Empirical Evaluation of Commercial Liberalism" (Ph.D. diss., Stanford University, 1997). For results that differ from ours, see Barbieri (fn. 24); and idem, "Economic Interdependence: A Path to Peace or a Source of Interstate Conflict?" *Journal of Peace Research* 33, no. 1 (1996). Our analyses to date indicate that this is primarily due to our different measures of interdependence: Barbieri does not weight trade by its contribution to GDP. The results reported in Oneal and Russett (fn. 6, 1999) show that the pacific benefits of trade, 1950–92, are robust to several alternative specifications, samples, and estimation procedures.

[55] Oneal and Russett (fn. 6, 1999).

are particularly prone to conflict. Then we set each continuous variab
at its mean value for the contiguous dyads, except that the lower de
pendence score was made equal to its median value, which is more rep
resentative. We postulated that the pair of states is not allied and do
not include a major power. We then estimated the annual probabili
that this "typical" dyad would be involved in a militarized dispute usin
the coefficients reported in columns 1 and 4 in Table 1. Next w
changed the theoretically interesting variables in succession by adding
standard deviation to the continuous measures or by making the dya
allied.

The first two columns of Table 2 give the percentage increase or de
crease in the annual risk of a dyad being involved in a dispute unde
these various conditions. Column 1 is based on the coefficients esti
mated using all dyads, and column 2 is produced with the coefficien
for just the politically relevant subset of cases.[56] Looking at the resul
in column 1, it is apparent that democracy and interdependence dra
matically reduce the likelihood of conflict. Compared with the typic
dyad, the risk that the more democratic dyad will become engaged in
dispute is reduced by 36 percent. If the dyad is more autocratic, th
danger of conflict is increased by 56 percent. A higher dyadic trade-to
GDP ratio cuts the incidence of conflict by 49 percent. A larger numbe
of joint memberships in IGOs has little effect on a dyad's likelihood o
conflict if all pairs of states are used in the estimation process. Whe
analysis is limited to the politically relevant dyads, however, the benef
of joint memberships in IGOs is clear. If the number of common mem
berships is fifty-three rather than thirty-two, the likelihood of conflic
is reduced by 13 percent. And when the analysis is limited to politicall
relevant pairs, the effects of democracy and economic interdependenc
are somewhat less than when all dyads are considered.

The substantive importance of the Kantian variables is confirmed i
their effects are compared with the results of changing the realist vari
ables. Consider again the second column of Table 2. If a state's prepon
derance of power is a standard deviation higher, that reduces th
probability of a dispute by 31 percent, but that result would require
fourfold increase in the capabilities of the stronger state. An allianc
lowers the incidence of interstate violence by 24 percent. This is sub
stantially less than when the dyad is more democratic or with a stan
dard deviation higher level of bilateral trade.

[56] This baseline probability is .031 among all dyads and .055 for the politically relevant pairs.

TABLE 2
PERCENTAGE OF CHANGE IN RISK FOR ANNUAL INVOLVEMENT IN A MILITARIZED DISPUTE FOR CONTIGUOUS DYADS[a]
(1886–1992)

	Based on		
	1. Equation 1 (All Dyads)	*2. Equation 1 (Politically Relevant Dyads)*	*3. Equation 2 (All Dyads)*
DEM_L increased by 1 std. dev.	−36	−33	
DEM_L decreased by 1 std. dev.	+56	+48	
DEPEND_L increased by 1 std. dev.	−49	−33	
IGO increased by 1 std. dev.	−2	−13	
CAPRATIO increased by 1 std. dev.	−27	−31	−33
ALLIES equals 1	−22	−24	−22
RELDEM_L increased by 1 std. dev.			−30
RELDEPEND_L increased by 1 std. dev.			−36
RELIGO increased by 1 std. dev.			−18
AVGDEM_L increased by 1 std. dev.			−26
AVGDEPEND_L increased by 1 std. dev.			−33
AVGIGO increased by 1 std. dev.			+3

[a] In each case other variables are held at baseline values.

We have argued that the characteristics of the less constrained state largely account for the likelihood of dyadic conflict, but the potential for violence may be significantly affected by the nature of the other dyadic member.[57] Democracies are more peaceful than autocracies at the national (or monadic) level as well as dyadically; but in our previous research we found, as Kant expected, that democracies and autocracies are particularly prone to fight one another because of the political distance separating them. Other analysts think that asymmetric interdependence may lead to conflict.[58] To evaluate these hypotheses we considered the influence of the higher democracy score and trade-to-GDP ratio, adding these variables to equation 1 both individually and as interactive terms with the lower democracy score or trade-to-GDP ratio.[59]

[57] Maoz (fn. 30); Oneal and Russett (fn. 6, 1997); Oneal and James Lee Ray, "New Tests of the Democratic Peace Controlling for Economic Interdependence, 1950–1985," *Political Research Quarterly* 50, no. 4 (1997).

[58] Robert O. Keohane and Joseph S. Nye, *Power and Interdependence: World Politics in Transition* (Boston: Little Brown, 1997); John A. Kroll, "The Complexity of Interdependence," *International Studies Quarterly* 37 (September 1993); Immanuel Wallerstein, "The Rise and Future Demise of the World Capitalist System," *Comparative Studies in Society and History* 16, no. 4 (1974); Barbieri (fnn. 24 and 54).

[59] If the effect of one variable (DEM_L, DEPEND_L) is thought to depend on the value of another (DEM_H, DEPEND_H), the test should include their interactive terms ($\text{DEM}_L * \text{DEM}_H$ and $\text{DEPEND}_L *$

The results, not reported in a table but available from the authors, indicated that the conflict-prone character of mixed pairs—one democracy and one autocracy—was limited to the post–World War II era. Plausibly the special institutional and ideological animosities between democrats and communists, solidified by the cold war, account for that. In the multipolar period, 1885–1939, dyads consisting of two democracies were the most peaceful after about 1900. Autocratic pairs and mixed dyads had similar rates of conflict. We found no evidence that asymmetric interdependence raised the likelihood of a militarized dispute. Increasing trade had significant pacific benefits whatever the relative size of the states involved. We did find a declining marginal utility for high levels of economic interdependence.[60]

DISENTANGLING THE SYSTEMIC AND CROSS-NATIONAL INFLUENCES OF THE KANTIAN MEASURES

Estimating equation 1 indicates that the likelihood of a dispute among all dyads is a function of the lower democracy score and the lower trade-to-GDP ratio in a dyad but not of states' joint memberships in international organizations. We suggested that the failure of the IGO variable to perform as expected results partly from including large numbers of irrelevant pairs of states that have no significant political relations and lack a realistic possibility of becoming engaged in a dispute. By contrast, limiting the analysis to contiguous dyads and those containing a major power highlights the benefits of international organizations. We also noted that our measure of joint IGO membership increases rather steadily over time. This may obscure the contribution of international organizations to peaceful interstate relations by making comparisons across time less meaningful, as with nominal GDPs in periods of inflation.

The influence of IGO membership can be reconsidered by distinguishing between the frequency of states' participation in international organizations through time and the standing of individual dyads relative to this annual measure at each point in time. We decompose each Kantian variable—the lower democracy score, the lower trade-to-GDP

DEPEND$_H$). See Robert J. Friedrich, "In Defense of Multiplicative Terms in Multiple Regression Equations," *American Journal of Political Science* 26, no. 4 (1982).

[60] Analyses in which we modeled the effect of interdependence as a hyperbola suggest that the benefits of trade increase rapidly and then approach a limit asymptotically. See Mark Gasiorowski and Solomon Polachek, "East-West Trade Linkages in the Era of Detente," *Journal of Conflict Resolution* 26, no. 4 (1982).

ratio, and the number of joint IGO memberships—into a systemic measure, the average value of states' democracy score, level of interdependence, or joint membership in IGOs (Figure 1), and a cross-sectional measure that ranks dyads relative to this annual average. The annual average of the number of joint IGO memberships (AVGIGO), for example, captures the prominence through time of international organizations, while the degree of involvement of individual dyads relative to this average (RELIGO) identifies those states that are more (or less) linked through the network of IGOs in each year.

To distinguish between the systemic and cross-sectional Kantian influences, we substitute in equation 1 AVGDEM and RELDEM_L for DEM_L, AVGDEPEND and RELDEPEND_L for DEPEND_L, and AVGIGO and RELIGO for IGO. Our second equation becomes:

$$\text{DISPUTE} = \text{RELDEM}_L + \text{RELDEPEND}_L + \text{RELIGO} + \text{AVGDEM} + \text{AVGDEPEND} + \text{AVGIGO} + \text{ALLIED} + \text{CAPRATIO} + \text{NONCONTIG} + \text{DISTANCE} + \text{MINORPWRS} \quad (2)$$

Column 1 of Table 3 reports the results of estimating equation 2 using all pairs of states. All the relative and systemic Kantian variables except the annual average of states' involvement in IGOs have a negative sign, indicating that increasing values reduce the likelihood of a militarized dispute; all but AVGIGO are very significant statistically. As explained in the last section, we standardized the three relative measures to permit direct comparison of their estimated coefficients. These indicate that economically important trade has the greatest conflict-reducing benefits, followed by democracy and joint memberships in international organizations. Two of the three Kantian systemic variables also affect the incidence of dyadic disputes: the likelihood of conflict drops when there are more democracies in the system and trade is more important economically; with both variables significant at the .001 level.[61] The influences of the other variables in the equation are

[61] There is a mild downward trend in the likelihood of a dispute over the period 1885–1992. To insure that the systemic Kantian variables were not simply collinear with this secular trend toward decreasing rates of disputes, we included in each of the equations reported in Table 3 an indicator of time, which equals the year minus 1884. The coefficients of the Kantian variables changed very little, and the average democracy score and trade-to-GDP ratio remained significant at the .001 level; the measure of time was never significant at the .05 level in these tests. If equation 2 is estimated for just the 1885–1939 period, the coefficient of the average level of interdependence becomes statistically insignificant, primarily because the level of trade at the outset of World War I was higher than it was during the interwar years; the average level of democracy remained significant at the .001 level.

TABLE 3
MODELS OF THE KANTIAN PEACE, 1886–1992:
PREDICTING INVOLVEMENT IN MILITARIZED DISPUTES
(DYADIC AND SYSTEMIC INFLUENCES, ALL DYADS)

Variable	*1. Only Kantian Systemic Variables*	*2. Systemic Kantian, Heg. Power, Satisfaction*	*3. Systemic Kanti Heg. Defense Bur*
Relative lower democ.	–0.3688***	–0.3576***	–0.4102***
(RELDEM$_L$)	(0.0680)	(0.0677)	(0.0703)
Relative trade/GDP	–0.7270***	–0.7045**	–0.5149**
(RELDEPEND$_L$)	(0.2333)	(0.2412)	(0.2132)
Relative IGO	–0.1304**	–0.1060*	–0.1602***
(RELIGO)	(0.0500)	(0.0512)	(0.0502)
Average democracy	–0.2383***	–0.2485***	–0.2702***
(AVGDEM)	(0.0412)	(0.0412)	(0.0423)
Average dependence	–292.4397***	–260.3094***	–355.5549***
(AVGDEPEND)	(36.4178)	(48.7066)	(39.7875)
Average IGOs	0.0043	0.0102	–0.0440***
(AVGIGO)	(0.0109)	(0.0115)	(0.0136)
Capability ratio	–0.2897***	–0.2787***	–0.3125***
(CAPRATIO)	0.0518	(0.0521)	(0.0135)
Alliances	–0.2554	–0.2186	–0.3330*
(ALLIES)	(0.1625)	(0.1665)	(0.1636)
Noncontiguity	–2.0080***	–2.0423***	–1.9225***
(NONCONTIG)	(0.1803)	(0.1828)	(0.1802)
Log distance	–0.4915***	–0.4637***	–0.5202***
(DISTANCE)	(0.0567)	(0.0597)	(0.0569)
Only minor powers	–2.0230***	–2.0073***	–2.0694***
(MINORPWRS)	(0.1893)	(0.1941)	(0.1911)
Hegemonic power		–1.5339	
(HEGPOWER)		(0.9502)	
Joint satisfaction		–0.0893	
(SATISFIED)		(0.1057)	
Heg. defense burden			17.9704***
(HEGDEF)			(1.9906)
Constant	–0.7345	–0.7113	–0.3735
	(0.4850)	(0.5075)	(0.4975)
Chi^2	1559.82	1530.24	1529.38
P of Chi^2	0.0000	0.0000	0.0000
N	149,372	147,963	149,372

*$p < .05$; **$p < .01$; ***$p < .001$, one-tailed tests

relatively unchanged. Preponderant power reduces the likelihood of a dispute, as do distance, an alliance, or the absence of a major power in the dyad. Using the onset of a dispute as the dependent variable produced nearly identical results.

The results of estimating equation 2 are important for three reasons. First, they show that dyads relatively more involved in international organizations at any point in time tend to be more peaceful, supporting the Kantian hypothesis regarding IGOs. Second, the results indicate that the statistical significance of democracy and the trade-to-GDP ratio in equation 1 is the consequence of temporal as well as cross-sectional variation. This is valuable assurance of the robustness of the pacific benefits of these Kantian influences. We now have explicit justification for believing that states can modify their circumstances by policies that increase democracy, interdependence, and, given the significance of the relative IGO measure, participation in international organizations. Third, it supports the view that there are systemic consequences of increasing democracy and trade for all pairs of states, not just for the liberal dyads.

The estimated coefficients for equation 2 allow us to compare the substantive importance of the relative and cross-sectional measures. We again calculate the probabilities of conflict for various hypothetical dyads. In calculating the baseline risk, we assume as before that the dyad is contiguous and set each continuous variable at its mean (or median for the trade ratio) for this subset of cases. We make the dyad unallied and assume it does not include a major power. We estimate the annual probability that this representative dyad would be involved in a dispute using the coefficients in column 1 of Table 3. Then one at a time we change each continuous variable by a standard deviation; finally we make the dyad allied.

Column 3 of Table 2 gives the annual probabilities of a dyad being involved in a dispute under these conditions. The effects of the cross-sectional Kantian variables, which rank dyads according to their position relative to the annual systemic averages, are again substantial. For dyads with a higher relative democracy score the risk of conflict is 30 percent below the baseline rate; a standard-deviation increase in relative dependence means a 36 percent lower probability of conflict; and when states' participation in IGOs is higher the likelihood of conflict is reduced by 18 percent. The substantive significance of the Kantian variables for interstate relations again emerges by comparing these effects with those that result from changing the realist variables.

A higher capability ratio means lowering the danger of violence by third, and when two states are allied the probability of conflict is low by 22 percent. Note also the effects of the Kantian systemic variable The risk of a dispute drops by 26 percent if the systemic average of th democracy score increases by a standard deviation (from –0.47 t +1.26); it falls 33 percent if the systemic average of the trade-to-GD ratio rises by a standard deviation (about 30 percent to .006). There effectively no change if the systemic average for states' participation i IGOs grows. Thus, two of the Kantian systemic variables have powerf effects throughout the international system. By normative or institu tional means, an increase in the number of liberal states constrains th use of force even by dyads that are not democratic or interdependent. The effect of IGOs is limited, however, to those states that participat jointly in more of these international forums relative to other pairs.

ASSESSING THE HEGEMON'S INFLUENCE ON DYADIC CONFLICT

In our last analyses we investigate the role of the hegemon. We firs evaluate a central claim of the theory of hegemonic stability an power-preponderance theory.[63] Both of these realist theories predic that conflict becomes more likely as the power of the leading state de clines relative to its principal rivals. At the same time, we also addres the argument that it has been the power of the (democratic) hegemo to reward its allies that accounts for the democratic peace. In a final tes we consider whether the hegemon's sense of its own insecurity, as indi cated by the ratio of its military expenditures to its gross domesti product, is associated with a heightened danger of conflict globally.

We assess the importance of the hegemon's relative power and state satisfaction with the status quo by adding two terms to equation 2 HEGPOWER, the proportion of the major powers' capabilities held b Britain (through 1939) and the U.S. (after 1945); and SATISFIED, ou measure of joint satisfaction, based on the similarity of each dyadi member's portfolio of allies to that of the leading power. It is appropri ate theoretically to include both in the same equation. If the hegemo is able to regulate the level of conflict in the international system, the

[62] To insure that the effects of the annual averages of the democracy score and trade ratio were tru systemic and not confined to only those dyads that were relatively democratic or interdependent, v added three interactive terms (AVGDEM*RELDEM$_L$, AVGDEPEND*RELDEPEND$_L$, and AVGIGO*RELIGO) equation 2. The results indicated that the effects of the systemic Kantian variables are not confined just those dyads that rank high relative to the annual averages.

[63] A. F. K. Organski, *World Politics* (New York: Knopf, 1968); George Modelski, ed., *Exploring Lo Cycles* (Boulder, Colo.: Lynne Rienner, 1987); Gilpin (fn. 43); Kugler and Lemke (fn. 39); K. Edwa Spiezio, "British Hegemony and Major Power War, 1815–1939: An Empirical Test of Gilpin's Mod of Hegemonic Governance," *International Studies Quarterly* 34, no. 2 (1990).

its influence should be greatest with those states with which it is most closely allied. At the same time the advantages for a state of aligning itself closely with the hegemon should be greatest when the power of the leading state is relatively large vis-à-vis its principal rivals; the hegemon in that situation should be most able to confer benefits upon its supporters.

Column 2 of Table 3 suggests that the strength of the leading state relative to its principal rivals does matter. The measure of hegemonic power is nearly significant ($p < .06$). Strong hegemony seems to reduce violence in the international system. This apparent effect stems, however, from the inability of a weakened hegemon (Britain) to prevent the outbreak of systemwide wars. In an analysis not reported in the table, the coefficient of our measure of hegmonic power reversed signs when the first year of each of the world wars was dropped: hegemony was then positively related to the incidence of disputes in the system ($p < .003$). Apparently the pacific benefits of hegemonic strength do not apply during normal periods of international relations. By contrast, we found no evidence in these analyses that states' satisfaction with the status quo accounts for the democratic peace. The measure of joint satisfaction in column 2 of Table 3 is far from statistical significance, while the significance of relative and systemic democracy is little changed.[64]

Finally we consider whether the hegemon's sense of its own security, as indicated by the proportion of GDP it devotes to military expenditures (HEGDEF), is related to the likelihood of dyadic conflict. We add our measure of the hegemon's defense burden to equation 2. The results of this test are reported in column 3. As seen there, the defense burden of the leading state is positively associated ($p < .001$) with the likelihood of dyadic disputes. There are wide-ranging consequences when the hegemon feels endangered. Nor is the heightened danger of conflict limited to the world wars, as with hegemonic power, or significant only for the hegemon or its allies. In a separate analysis not reported in the table, we confirmed that other states, too, experience more disputes when the hegemon has increased the proportion of its resources committed to the military. Our systemic and relative Kantian variables nonetheless remain important. Even the systemic measure of states' participation in international organizations is now significant at the .001 level. The effectiveness of IGOs may depend in part upon the

[64] We tested alternative specifications in evaluating the role played by states' satisfaction with the status quo. We adopted the weak-link assumption, adding the smaller of the tau-b measures of satisfaction to equation 2, and investigated whether two dissatisfied states might also be peaceful; but these terms were not statistically significant.

major powers not feeling a need to develop, and presumably use, independent military means for protecting and promoting their interests.

A KANTIAN SYSTEM? PAST AND FUTURE

Our analyses for the years 1885–1992 indicate that Kant was substantially correct: democracy, economic interdependence, and involvement in international organizations reduce the incidence of militarized interstate disputes. The pacific benefits of the Kantian influences, especially of democracy and trade, were not confined to the cold war era but extend both forward from that era and back many decades. Moreover, these benefits are substantial. When the democracy score of the less democratic state in a dyad is higher by a standard deviation, the likelihood of conflict is more than one-third below the baseline rate among all dyads in the system; a higher bilateral trade-to-GDP ratio means that the risk of conflict is lower by half. The pacific benefits of democracy in the twentieth century are clear, and the change from the nineteenth century is consistent with an evolutionary view: democratic institutions matured, and the suffrage was extended. In addition, as Kant believed, states may learn from the success and failure of their policies.

The benefits of joint membership in intergovernmental organizations are more modest but nevertheless significant for the politically relevant dyads—contiguous states and dyads containing at least one major power. For these particularly dangerous dyads, the probability of a dispute drops by 13 percent when the number of joint memberships in IGOs is greater by a standard deviation. The pacific benefits of international organizations are also apparent when the trend in this variable is eliminated: among all dyads, pairs of states more involved by a standard deviation in IGOs relative to the annual systemic average are 18 percent less likely to become embroiled in interstate violence.

By distinguishing the influences of the Kantian systemic averages from the standings of each dyad relative to the annual means, we also showed benefits of democracy and trade over time as well as cross-sectionally. The effects of the systemic Kantian influences on dyadic conflict are important. The international system is more peaceful when there are more democracies and when trade is greater. *All* dyads—even those not democratic or interdependent—become less dispute-prone when those systemic Kantian variables increase. The constraining effect of norms and institutions that emerge when there are more democracies and when trade is economically important for many states holds

even for those that participate to only a limited degree in the Kantian subsystem.[65]

Over the period 1885–1992 states' participation in IGOs rose steadily, but there is little evidence of a trend toward increased democracy or economic interdependence over the complete span of time. A long trend toward greater interdependence may be masked by two aspects of our data. First, the sample changes over time. Less developed and more peripheral states are probably underrepresented before World War I. Only with the establishment of the IMF and UN agencies does information on states' wealth and dyadic trade become reasonably complete. Thus, the average level of bilateral interdependence may be overstated in the early years. Second, decolonization in the late 1950s and the 1960s created dozens of new states that were less democratic and less integrated into the global economy than the states already in the system, lowering the average scores for democracy and interdependence. And as noted, the codings of democracy that we use overstate the democratic character of states in much of the nineteenth century before suffrage was extended to women and those without property.

Both democracy and interdependence do show a marked jump after World War II. The number of democracies has grown steadily since the late 1970s, especially after the cold war ended. Trade grew rapidly in the 1970s. Since 1987 these phenomena have been followed by a precipitous drop in the number of interstate wars, despite the entry of many new states into the system.[66] Our results for the early post–cold war years cover only 1989–92, but they indicate that the beneficial effects of democracy, interdependence, and IGOs continued past the end of the cold war. Moreover, our analyses of the 1885–1992 period suggest that the relative peace of the past decade owes less to the systemic effects of power and hegemony than to growing Kantian influences.

As for the realist influences, some of the dyadic characteristics—chiefly distance, power preponderance, and minor power status—also reduce the likelihood of disputes. This is not surprising, though the lack of a robust effect for alliances is. The Kantian influences have not abolished power politics. Realist variables at the systemic level also make a difference in the incidence of dyadic conflict. Both world wars occurred when Britain, the hegemonic state, was weak. Yet hegemony does not always work as hypothesized. During more normal periods of

[65] See the references in fnn. 17 and 19 and the textual discussion accompanying them.
[66] Monty G. Marshall, *Third World War* (Lanham, Md.: Rowman, Littlefield, 1999).

international relations, there were more militarized disputes when the hegemon was powerful than when it was weak; and when the hegemon felt threatened (as evidenced by higher military spending relative to its gross domestic product), the likelihood of disputes rose throughout the system.

Democracies fought two world wars side by side, along with some autocracies that shared their strategic interests. Was the democracies' common alignment purely a result of strategic interests? It is more likely that shared interests in democracy and economic freedom played an important role. By contrast, alliances had no systematic dispute-inhibiting influence prior to the cold war. For the post-1945 era, when a reasonably strong effect of alliances is evident, it strains belief to attribute that effect primarily to strategic interests. Of course the cold war was substantially about national security as understood by realists. But it was also about a clash of two fundamentally different political and economic systems. The governments, dominant classes, and peoples of the free-market democracies felt not only that their physical security and national independence were threatened but also that their prosperity and especially their political and economic liberties were at stake. Hence they allied with one another to preserve their common way of life.[67]

The post–cold war era is full of affirmations about the importance of democracy, freedom, and prosperity built on interdependent markets. Some may be just rhetoric, but sophisticated global economic actors understand the role that interdependence plays in their prosperity. In 1999 NATO fought a war against Serbia in the name of democracy and human rights in Europe, against a dictatorial government that did not constitute a strategic threat. In time we shall see whether peace will hold among democracies and interdependent states, but to call the democratic peace "a byproduct of a now extinct period in world politics"[68] sounds very like a premature report of its death.

Analytically, we are progressing toward a synthesis of Kantian and realist influences and of dyadic and systemic perspectives. Kant argued that three naturally occurring tendencies operate to produce a more peaceful world. Individuals desire to be free and prosperous, so democracy and trade will expand, which leads to the growth of international law and organizations to facilitate these processes. Peace, therefore,

[67] By controlling for states' interests, we have tried to show that the democratic peace is not an artifact of the cold war; see Oneal and Russett, "Is the Liberal Peace Just an Artifact of Cold War Interests? Assessing Recent Critiques," *International Interactions* 25, no. 3 (1999).

[68] Gowa (fn. 3), 114.

does not depend upon a moral transformation of humanity as long as even devils are self-interested and can calculate.[69] For Kant, a child of the Enlightenment, this was evidence of an ordered universe and, perhaps, of providential design. Yet he did not think that the process was mechanical or the outcome certain: reason would not always prevail, and states and individuals would not always act in conformity with their enlightened interests. Human agents must learn from experience, including that of war, and change behavior.

The current unipolar character—inevitably transitory—of our world, with no other state close to the power of the United States, provides an opportunity to build a peace based not only on military force but also on Kantian principles. Hegemony does not last forever. Consequently, democracy should be extended and deepened, the "cosmopolitan law" of commerce expanded, and international law and respect for human rights institutionalized. Kant would say this is a moral imperative.

APPENDIX: VARIABLES

ALLIES: 1 if dyad members linked by defense treaty, neutrality pact, or entente
AVGDEM: average democracy score for all states in a year
AVGDEPEND: average dyadic trade to GDP ratio for all states in a year
AVGIGO: average number of dyadic shared IGO memberships
CAPRATIO: logarithm of ratio of higher to lower power capability in a dyad
DEM_H: higher democracy score in a dyad
DEM_L: lower democracy score in a dyad
DEPEND_H: higher dyadic trade-to-GDP ratio in a dyad
DEPEND_L: lower dyadic trade-to-GDP ratio in a dyad
DISPUTE: involvement in dyadic dispute
DISTANCE: logarithm of dyadic distance in miles between capitals or major ports
HEGDEF: ratio of leading state's military spending to its GDP
HEGPOWER: leading state's proportion of the capabilities of all major powers
IGO: number of international organization memberships shared by a dyad
MINORPWRS: 1 if dyad does not include a major power
NONCONTIG: 1 if dyad is not contiguous by land border or less than 150 miles of water
RELDEM_L: DEM_L – AVGDEM/standard deviation of DEM
RELDEPEND_L: DEPEND_L – AVGDEPEND/standard deviation of DEPEND_L
RELIGO: IGO – AVGIGO/standard deviation of IGO
SATISFIED: tau-b measure of similarity of dyad members' alliance portfolios to that of the leading state

[69] Kant (fn. 1), 112.

THE HIGH POLITICS OF IMF LENDING

By STROM C. THACKER*

INTRODUCTION

CONSIDERING the degree of scrutiny given to the role of the International Monetary Fund (IMF, or Fund) in the international economy, we know little about the underlying causes of the IMF's behavior.[1] During the 1980s, the IMF became a "lender of last resort" for many developing country governments that had been cut off from private credit markets and faced destabilizing imbalances of payments. After private capital began to return voluntarily to what were called the emerging markets in the early 1990s, the anticipated erosion of the Fund's role in the developing world did not materialize. Faced with recurrent payments' imbalances, pressures for currency devaluation, and the macroeconomic instability associated with crises in Latin America, Asia, and Russia, the developing nations have turned with increasing frequency to IMF credits and stabilization plans. Despite the growing body of research on the IMF's critical role in international finance, we still have few explanations of and only patchy empirical data on why the IMF approves loans to some countries but not others. As the Fund delves further into the management of balance of payments and currency crises around the world, both theoretical and practical imperatives dictate that we specify more fully and test more systematically

*A preliminary version of this paper was presented at the 1997 annual meeting of the American Political Science Association in Washington, D.C. I would like to thank Tim McKeown, Patrick Conway, William Bernhard, Mary Matthews, Dane Rowlands, the participants in the faculty research seminar in the Department of International Relations at Boston University, and the anonymous *World Politics* referees for their insightful comments and suggestions. Yvonne Ochoa and Jaya Badiga provided able research assistance. The usual disclaimers apply.

[1] The literature on the IMF is extensive. For useful surveys, see Graham Bird, "The International Monetary Fund and the Developing Countries: A Review of the Evidence and Policy Options," *International Organization* 50, no. 3 (1996); idem, *IMF Lending to Developing Countries: Issues and Evidence* (London: Routledge, 1995); Sebastian Edwards, "The International Monetary Fund and the Developing Countries: A Critical Evaluation," NBER Working Paper, no. 2909 (1989); Tony Killick, *IMF Programs in Developing Countries: Design and Impact* (London: Routledge, 1995); John Williamson, ed., *The Lending Practices of the International Monetary Fund* (Washington, D.C.: Institute for International Economics, 1982); and idem, *IMF Conditionality* (Washington, D.C.: Institute for International Economics, 1983).

competing explanations of IMF behavior. What factors influence the IMF's decisions to lend? Are these decisions based on technical economic criteria, or do they reflect the political preferences of the Fund's more powerful members? What are those preferences, and how do they affect the IMF's relationship with its developing country clients? In other words, does politics matter? More importantly, *how* does politics matter?

The literature on international institutions and multilateralism suggests that the operation of multilateral economic organizations like the IMF will assume growing importance in a posthegemonic international order.[2] This body of theory raises several important questions: To what extent do multilateral institutions modify the interests or constrain the behavior of their member states? Can the more power ful states use these organizations as effective instruments of national foreign policy, or are such pressures diluted or transformed in passing through multilateral channels? Finally, what underlying political interests drive the behavior of the large powers within the multilateral institutions, and how do they do so? On the practical side, Gallarotti has shown that poorly managed international organization not only can be ineffective but also can destabilize the international system.[3] The debates surrounding proposed increases in Fund resources and recent loan packages negotiated with South Korea, Indonesia, Russia, and Brazil demonstrate the growing popular recognition of these kinds of problems, but scholarly research has yet to address these questions adequately.

Economists have made inroads in isolating the economic bases of IMF lending, but they are the first to point out that their models remain incomplete. Researchers have found statistically significant results for

[2] For effective treatments of these and related issues, see Robert O. Keohane, *After Hegemony: Cooperation and Discord in the World Political Economy* (Princeton: Princeton University Press, 1984); idem, *International Institutions and State Power: Essays in International Relations Theory* (Boulder, Colo.: Westview, 1989); idem, "Multilateralism: An Agenda for Research," *International Journal* 45, no. 1 (1990); John Gerard Ruggie, ed., *Multilateralism Matters: The Theory and Praxis of an Institutional Form* (New York: Columbia University Press, 1993); Stephen D. Krasner, ed., *International Regimes* (Ithaca, N.Y.: Cornell University Press, 1983); and Kenneth A. Oye, ed., *Cooperation under Anarchy* (Princeton: Princeton University Press, 1986).

International institutions and multilateralism are not necessarily equivalent. The IMF fits Ruggie's definition in *Multilateralism Matters* of multilateral organization as "defined by such generalized decision-making rules as voting or consensus procedures" (p. 14). On IMF decision-making procedures, see Frank Southard, "The Evolution of the International Monetary Fund," Princeton Essays in International Finance, no. 135 (1979); and Frederick K. Lister, *Decision-Making Strategies for International Organizations*, vol. 20, *World Affairs* (Denver, Colo.: Graduate School of International Studies, University of Denver, 1984).

[3] Giulio M. Gallarotti, "The Limits of International Organization: Systematic Failure in the Management of International Relations," *International Organization* 45, no. 2 (1991).

[illegible] impact of a number of economic variables on IMF lending, bu
low overall explanatory power of the econometric models reviewe
Bird suggests that "there are probably a range of other non-econc
factors which still need to be delineated."[4] One likely source of none
nomic factors is politics, but political scientists have not yet been abl
demonstrate the systematic impact of political variables on IMF ler
ing.[5] Several case studies offer suggestive, but not generalizable, ev
dence of the political bases of IMF lending. Fewer studies ha
attempted to construct a systematic political explanation of Fund be
havior. This paper attempts to fill some of those gaps in the literatur
and proposes answers to those questions by developing and testing sta
tistically a political explanation of IMF lending patterns.

The IMF's stated decision-making procedures prohibit the consideration of political factors. Loans are made strictly on the basis of the monetarist "Financial Programming" model and a "Doctrine of Economic Neutrality" that is blind to such factors as international politics and the nature of developing country regimes.[6] The Fund may impose strict lending requirements, but it applies them fairly to all countries. The meetings of the IMF executive board, which approves all Fund programs, are highly secretive.[7] The specific considerations that determine the board's deliberations and decisions are therefore not available in the public domain. For its part, the Fund staff publicly maintains the position of economic neutrality, but evidence presented in numerous case studies leaves open the possibility that political factors play an important role.

There are at least three reasons to suspect that politics matters in the IMF. First, several studies have found extremely low rates of borrower compliance with Fund conditionality, yet the IMF continues to lend to many of these problem debtors even after earlier programs have been canceled for noncompliance.[8] Finch, a former IMF staff member, suggested in the late 1980s that economic factors could not explain these

[4] Bird (fn. 1, 1995), 124.

[5] Dane Rowlands, "Political and Economic Determinants of IMF Conditional Credit Arrangements: 1973–1989" (Manuscript, Norman Paterson School of International Affairs, Carleton University, Ottawa, Ont., 1995).

[6] Jacques J. Polak, "Monetary Analysis of Income Formation and Payments Problems," IMF Staff Papers, no. 6 (1957), cited in Edwards (fn. 1); and idem, "The Changing Nature of IMF Conditionality," Princeton Essays in International Finance, no. 184 (1991); and Richard Swedberg, "The Doctrine of Economic Neutrality of the IMF and the World Bank," *Journal of Peace Research* 23 no. 4 (1986).

[7] R. S. Eckaus, "How the IMF Lives with Its Conditionality," *Policy Sciences* 19 (October 1986).

[8] Southard (fn. 2), 13; Edwards (fn. 1); C. David Finch, "The IMF: The Record and the Prospects," Princeton Essays in International Finance, no. 175 (1989); and John Spraos, "IMF Conditionality: Ineffectual, Inefficient, Mistargeted," Princeton Essays in International Finance, no. 166 (1986).

as: "Because decisions were no longer based on compatibility epayment terms, lending was guided increasingly by the political rences of the leading industrial countries."[9] Second, each country's esentative on the Fund's executive board is appointed by his or her he government (Treasury, in the case of the United States). Thus it uld come as no surprise that the positions of those representatives thin the Fund reflect the political interests of the national governents they serve.[10] As Smith puts it, "The IMF is itself a political institution. It is managed by politically appointed individuals from member ations, and the political interests of its members influence its decisions."[11] Although the staff is less directly linked to national governments, the executive board must approve all proposed programs. The familiarity of Fund staff members with the preferences of the executive board discourages them from submitting loan packages that the board is likely to veto.[12]

Finally, weighted voting and the decision-making procedures of the Fund also leave room for politics. As of April 1995, the U.S. controlled 17.83 percent of the voting power in the IMF, followed by Germany and Japan with 5.5 percent each, and France and the United Kingdom with 5.0 percent each.[13] An evolving system of special majorities has helped the U.S. maintain its influence beyond that dictated by its gradually decreasing voting share.[14] An 85 percent majority is required for the most important Fund decisions, giving the U.S. alone, and other groups of countries together, veto power. The U.S. can also push through favored programs, which it might not be able to do based on its votes alone. Although the managing director has traditionally been a European, he rarely acts against U.S. preferences.[15] That is not surprising since he is appointed through a process over which the U.S. has veto power.[16] But

[9] Finch (fn. 8), 2.

[10] Lars Schoultz, "Politics, Economics, and U.S. Participation in Multilateral Development Banks," *International Organization* 36, no. 3 (1982); Benjamin J. Cohen, "International Debt and Linkage Strategies: Some Foreign Policy Implications for the United States," in Miles Kahler, ed., *The Politics of International Debt* (Ithaca, N.Y.: Cornell University Press, 1986).

[11] Fred L. Smith, "The Politics of IMF Lending," *Cato Journal* 4 (Spring/Summer 1984). The U.S. representative is "ordered by law to clear his or her decisions with the Secretary of the Treasury." Swedberg (fn. 6), 379.

[12] Kendall W. Stiles, *Negotiating Debt: The IMF Lending Process* (Boulder, Colo.: Westview, 1991).

[13] *IMF Annual Report* (Washington, D.C.: IMF, 1995), 216.

[14] See Lister (fn. 2).

[15] Samuel Lichtensztejn and Mónica Baer, *Fondo Monetario Internacional y Banco Mundial: Estrategias y Políticas del Poder Financiero* (Mexico City: Ediciones de Cultura Popular, 1987), 60–61.

[16] Miles Kahler notes that the U.S. has in the past refused to support a renewal of the managing director's tenure when his "accomplishments did not meet American expectations." Kahler, "The United States and the International Monetary Fund," in Margaret P. Karns and Karen A. Mingst, eds., *The United States and Multilateral Institutions* (Boston: Unwin Hyman, 1990), 94.

that power rarely needs to be wielded openly. The managing dire typically makes decisions based on a "sense of the meeting," deri from the comments of the various participants and their relative vot power.[17] Other powers can be reluctant to speak against the U.S. fear that the U.S. will later retaliate by exercising its veto power or their own favored programs.[18] Finally, the U.S. and its like-minded lies together can effect international monetary cooperation by formir subsets, or "k-groups," of countries to push through certain package that single parties cannot.[19]

Given the possibility that political factors influence IMF decisions several scholars have argued that the Fund's more powerful members manipulate it to further their own political and economic interests.[20] The U.S. government, for its part, "has repeatedly told foreign governments that it will not intervene in negotiations between the Fund and member governments."[21] Kahler notes, however, that "the U.S. (and other major countries) can still influence programs for friends and clients at the margin."[22] Others suggest that American politicization of Fund lending is more widespread. A series of case studies conducted for a project directed by Killick and Bird reveals that at least one-third of the seventeen countries studied secured favorable loan terms on their IMF programs due to the intervention of major shareholding countries on their behalf.[23] Stiles concludes that in only one of seven cases examined did the Fund adopt a politically neutral, technocratic approach to lending.[24]

Such case studies have been useful for providing the kind of rich detailed data that are unavailable by other means, for formulating testable hypotheses, and for providing initial evidence of the role of politics in IMF lending. Despite these advances, we are still unable to say much more than that politics *seems* to matter, at least in some cases. This

[17] The origins of this procedure date back to the Fund's early years, when the U.S. executive director went to great lengths to muffle the strong voice of U.S. power, which nevertheless was decisive. See Southard (fn. 2), 5–6, 19–20.

[18] Eckaus (fn. 7), 237; Stiles (fn. 12), 37.

[19] Ruggie (fn. 2), chap. 1; James A. Caporaso and Miles Kahler attribute part of the postwar economic cooperation to this type of "minilateralism." The creation of the Bretton Woods monetary order through U.S. and British coordination and the subsequent adjustments made by the G-7 after its breakdown (for example, the Plaza and Louvre accords) can be profitably understood in these terms. Caporaso, "International Relations Theory and Multilateralism: The Search for Foundations," and Kahler, "Multilateralism with Small and Large Numbers," in Ruggie (fn. 2).

[20] Cheryl Payer, *The Debt Trap: The International Monetary Fund and the Third World* (New York: Monthly Review Press, 1974); and Swedberg (fn. 6).

[21] Kahler (fn. 16), 110.

[22] Ibid.

[23] Killick (fn. 1), 118–19.

[24] Stiles (fn. 12), 196–97.

r aims to accomplish two essential tasks. First, it attempts to pro- the first systematic quantitative evidence for whether politics af- s IMF behavior. Second, it proposes a dynamic explanation of how itical factors affect interactions of multilateral organizations with ir member states, and tests the statistical formulation of that argu- ent in the case of the IMF. I first propose a simple macroeconomic odel and a political explanation of IMF lending. I then operationalize hese hypotheses together and report the results of a series of statistical sts conducted on them jointly. I conclude with a discussion of the imitations and broader implications of this research.

A Simple Macroeconomic Model

Economists have isolated several important demand- and supply-side macroeconomic determinants of IMF lending. Rather than attempt to replicate such studies, I take them as a starting point for my political analysis. Conway has modeled participation in IMF programs as a function of a country's economic environment and its past economic performance.[25] He finds statistically significant negative results for lagged variables representing the ratio of foreign exchange reserves to imports, the growth rate of real gross national product, the ratio of the current account to GNP, the terms of trade, and the real rate of interest.[26] Variables capturing level of development (proxied by the share of output from the agricultural sector) and long- and short-term debt stocks did not attain conventional levels of statistical significance.

Lindert tests the impact of several variables on the interest rate charged on official creditor lending to fifty-one countries in 1985.[27] He obtained statistically significant results for only two variables—the log of absolute nominal public and publicly guaranteed debt in 1981 and the log of per capita income—each with a positive coefficient. None of the other variables—the share of debt held by official creditors, the ratio of debt service to GNP, the ratio of reserves to imports, money stock growth, prior default, prior rescheduling, and years since first rescheduling—approached conventional levels of statistical significance.

Summarizing his own and others' research, Bird identifies statistically significant regression results with negative coefficients for balance

[25] Patrick Conway, "IMF Lending Programs: Participation and Impact," *Journal of Development Economics* 45, no. 2 (1994).

[26] He finds statistically significant positive results for prior participation and the percentage of available funds drawn down. A series of dummy variables for each year had generally significant results.

[27] Peter H. Lindert, "Response to Debt Crisis: What Is Different about the 1980s?" in Barry J. Eichengreen and Lindert, eds., *The International Debt Crisis in Historical Perspective* (Cambridge: MIT Press, 1989).

[illegible] and reserves. [illegible] significant positive regi[illegible] [illegible] to private bank credit,[29] domestic [illegible] spending. With the exception of Conwa[illegible] [illegible] power of these models suggests that they are under[illegible]ified. Rather than attempt to test the validity of distinct comp[illegible] macroeconomic models, this paper draws upon the large body of e[illegible]ing research to identify the putative economic determinants of F[illegible] activities. First, the balance of payments position of a country is the [illegible]tial baseline upon which its participation in IMF programs is evaluat[illegible] An improvement in the balance of payments is the stated primary g[illegible] of most IMF lending programs,[30] and without a payments deficit, [illegible] country should neither need nor be eligible for Fund lending.[31] Wh[illegible] faced with a payments deficit, a country can either run down its re[illegible]serves or borrow internationally.[32] In the context of the debt crisis, run[illegible]ning down reserves was not a viable long-term solution, and the mos[illegible] reliable source of international borrowing was the IMF. Specifically, de[illegible]terioration in the balance of payments is expected to increase th[illegible] chances of receiving a loan from the IMF.

Second, a country's debt position should affect its demand for and the supply of an IMF loan for distinct reasons. On the demand side, a heavier debt burden increases developing countries' need for external finance to service that debt. In terms of supply, some have argued that the more heavily indebted countries have more bargaining leverage over the IMF because of their importance to global financial stability.[33] In addition, some view IMF loans as a payoff to foreign creditors.[34] Assuming that those lenders wield influence within the executive board, IMF loans more likely will go to countries where creditors are more heavily exposed. Lindert found that the more heavily indebted countries did receive official creditor loans, but at higher interest rates than the smaller debtors, due to the reluctance of Northern taxpayers to finance more

[28] Bird (fn. 1, 1995).

[29] This result supports the catalytic impact of IMF lending as providing a "seal of approval" that encourages private banks to resume lending to a country that has negotiated an agreement with the Fund. Bird (fn. 1, 1995), 122. A negative result would suggest a substitution effect between IMF and private lending.

[30] Spraos (fn. 8); and Finch (fn. 8).

[31] Bird (fn. 1, 1995), 109.

[32] Ibid., 23.

[33] See Jahangir Amuzegar, "The IMF under Fire," *Foreign Policy* 64 (Fall 1986).

[34] See Walden Bello and David Kinley, "The IMF: An Analysis of the International Monetary Fund's Role in the Third World Debt Crisis, Its Relation to Big Banks, and the Forces Influencing Its Decisions," *Multinational Monitor* 4 (1983).

ry terms.[35] For the purposes of this study, the fact that the ors were more likely to receive loans is of primary interest.
the level of per capita income of a country also may influence to secure Fund assistance. Killick notes that the IMF's historrow focus on balance-of-payments considerations has given recent years to a broader view that acknowledges the relationtween the balance of payments and growth.[36] Lindert reports ficial favoritism for poor countries resulted in lower interest rates ficial loans, and Bird argues that poorer countries are less likely to ow on private capital markets and therefore to have a higher reladmand for IMF loans.[37] Countries with lower per capita incomes ld be more likely to request and receive a loan from the IMF.
Fourth, if the Doctrine of Economic Neutrality is followed, a poor dit history should decrease the chances of receiving a loan. After the ssive defaults of the 1930s, the debtor nations were effectively cut off om credit for several years. Many have argued that we should expect milar outcomes now.[38] Specifically, past failures to uphold IMF loan reuirements should make it more difficult to receive additional loans.
Finally, both neomarxist and modern political economy interpretaions would suggest two additional macroeconomic indicators that hould affect Fund decisions due to the influence of "low politics": the rade and investment exposure of firms based in the IMF's major princiial shareholder, the U.S. Authors grouped loosely within a neomarxist or dependency) perspective argue that capitalists in the core states, esecially the U.S., dictate IMF policy at the expense of the nations of the eriphery.[39] A political economist more concerned with the impact of lomestic politics on foreign economic policy might also posit, without ecessarily adopting the concomitant exploitation argument, that wellrganized export enterprises and multinational corporations (MNCs) ressure the U.S. government to protect their interests on the executive oard. Neomarxism suggests that higher levels of U.S. exposure lead to greater likelihood of receiving an IMF loan because the attendant pol-

[35] Lindert (fn. 27), 245.
[36] Killick (fn. 1).
[37] Lindert (fn. 27), 243; Bird (fn. 1, 1995), 112.
[38] Barry Eichengreen has questioned the impact of the "default penalty" on future credit access. ichengreen, "The U.S. Capital Market and Foreign Lending, 1920–1955," in Jeffrey D. Sachs, ed., *eveloping Country Debt and the World Economy* (Chicago: University of Chicago Press, 1989), 247. f. Jonathan Eaton and Mark Gersovitz, "Debt with Potential Repudiation: Theoretical and Empiril Analysis," *Review of Economic Studies* 48 (April 1981).
[39] See E. A. Brett, "The World's View of the IMF," in Latin America Bureau, ed., *The Poverty Brors: The IMF and Latin America* (London: Latin America Bureau, 1983); Manuel Pastor, "The Effects IMF Programs in the Third World: Debate and Evidence from Latin America," *World Development* (Fall 1987); and Swedberg (fn. 6).

icy conditionality promotes the expansion of global capitalism. A less explored, domestic-politics interpretation yields more ambiguous expectations. The inflow of foreign exchange and the restoration of international creditworthiness would be expected to benefit U.S. exporters and foreign investors, while the demand-reduction components of the typical IMF program would suggest a negative impact for these variables.[40]

HIGH POLITICS AND THE IMF

The international political aspects of IMF lending have received far less rigorous analysis. Two rudimentary strands of thought comprise this genre, but neither has been fully developed nor adequately tested. I label the first, more common strand the "political proximity" hypothesis. Simply put, political friends of the U.S. are more likely to receive loans than are its enemies. In addition to the case studies described above, Bello and Kinley argue that the U.S. disregarded the Fund's economic criteria and pressured the Fund to approve loans to politically friendly South Africa, El Salvador, and Haiti.[41] The IMF has also denied loans to economically worthy political enemies of the U.S., such as Vietnam.[42] In sum, the more closely a country aligns with the U.S., the higher the probability it will receive a loan from the IMF.

These arguments have not yet been fully developed conceptually nor thoroughly tested empirically. To illustrate an intuitive analytical foundation for this argument and to facilitate its testing, I construct a continuous voting space, scaled from 0 to 1, where 1 represents total agreement and 0 complete discord with the United States on a single broad dimension of foreign policy affinity (such affinity could be easily measured by votes in a majority-rule international voting arena, such as the United Nations General Assembly). Figure 1 is a graphic representation of this space and a schematic portrayal of the political proximity hypothesis. Countries at point A, at the far left-hand side of the voting space, have little chance of receiving a loan, while those at point C, at the far right, are much more likely to receive Fund support. Alignment near the middle, at point B, has little or no effect.

[40] The net effect of DFI exposure may depend on the sectoral location of the investment. If it serves primarily the domestic market, a negative result might be expected. If it serves mostly export markets, a positive result would be more likely. The impact of export exposure may depend on whether the product exported is a final consumption good (negative) or an input into the export sector (positive).

[41] Bello and Kinley (fn. 34), 14.

[42] Susumu Awanohara, "Fiscal Interdiction: U.S., Japan Block IMF Effort to Support Vietnam," *Far Eastern Economic Review*, September 28, 1989.
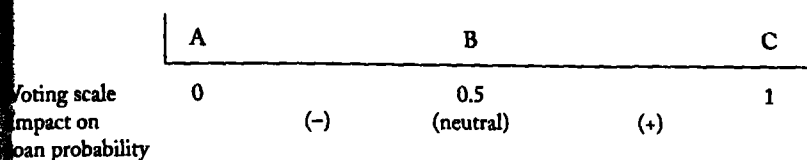
FIGURE 1
THE POLITICAL PROXIMITY HYPOTHESIS

Despite the existence of plentiful case studies, previous research has ot effectively evaluated this argument. Furthermore, several studies ave documented numerous cases where U.S. "enemies" are rewarded r "friends" punished. In many instances the Fund has made loans to eftist governments, such as Manley's in Jamaica in 1979 and the East uropean Soviet bloc countries of Hungary, Yugoslavia, and Romania, ach of which ranked among the top fifteen IMF loan recipients from 952 to 1984.[43] This evidence seems to paint a picture of IMF lending as n apolitical technocratic process, economic neutrality at its best. But hese loans may not have been justifiable on purely economic grounds, ither. Assetto compares the results of a regression equation designed o predict the size of IMF loans based solely on economic criteria with he size of the actual loans received by the three East European counies to conclude that actual lending exceeded predicted lending by a gnificant margin.[44]

A less static interpretation of these anomalies introduces the "politial movement" hypothesis, the less-developed second strand of the potical argument. Movement toward or away from the U.S. on nternational political issues may be at least as important as the absolute olitical alignment of a particular country. Hinting at this idea, lorowitz asks whether the IMF should use loans to entice countries like omania and Hungary away from the Soviet bloc.[45] This notion is conistent with the cases of Hungary, Yugoslavia, and Romania, all of hom distanced themselves politically from the Soviet Union (that is, oved closer to the U.S.). In contrast, neither Czechoslovakia nor

[43] Amuzegar (fn. 33); Valerie J. Assetto, *The Soviet Bloc in the IMF and the IBRD* (Boulder, Colo.: estview, 1988).

[44] Ibid., 50.

[45] Irving Louis Horowitz, "The 'Rashomon Effect': Ideological Proclivities and Political Dilemmas the IMF," in Robert J. Myers, ed., *The Political Morality of the International Monetary Fund* (New unswick, N.J.: Transaction Books, 1987), 96.
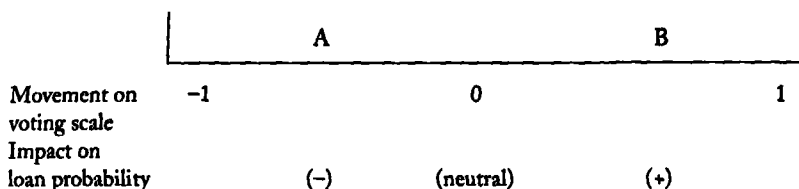
FIGURE 2
THE POLITICAL MOVEMENT HYPOTHESIS

Poland, more consistently faithful Soviet allies, received any IMF funding during this period.[46]

Frey applies Hirschman's neutrality model to formulate a model of the bilateral aid-giving process in a bipolar world where aid recipients can play the two superpower donors off one another.[47] On a more general level, McKeown models the aid relationship formally as a sequential bargaining game between the lending principal and borrower.[48] The lender exchanges aid for political realignment by a developing country toward the position of the lender. The borrower moves from its "ideal" point to a new equilibrium point where the marginal utility of additional aid received equals the domestic political loss incurred by another move away from its ideal position. I adapt and extend these central insights to hypothesize that political movement toward the U.S. increases a country's probability of receiving a loan from the IMF.

I portray the lending process as a dynamic game between each borrower and a single lender. I do not model this interaction in formal game-theoretic terms, nor do I model the relationship between the U.S. and IMF. Rather, I assume the U.S. plays the role of principal within the IMF, generate testable hypotheses about the relationship between the Fund and borrowing countries, and evaluate them empirically. If the data confirm these hypotheses, it would strongly suggest, but not directly confirm, a predominant U.S. presence in an increasingly important multilateral organization. Such a conclusion would have important implications for the study of international institutions and regimes, as well as for the multilateral management of the international economy.

[46] Assetto (fn. 43), 184.

[47] Bruno S. Frey, *International Political Economics* (London: Basil Blackwell, 1984), chap. 5; Albert O. Hirschman, "The Stability of Neutralism: A Geometrical Note," *American Economic Review* 54 (March 1964).

[48] Timothy J. McKeown, "Resolving the 'Conditionality Paradox' in U.S. Bilateral Foreign Aid" (Manuscript, University of North Carolina, Chapel Hill, n.d.).
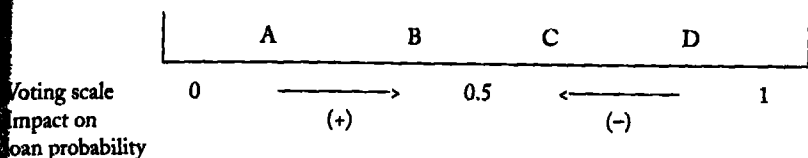
FIGURE 3
THE IMPACT OF POLITICAL MOVEMENT OVER TIME

Figure 2 captures the basic argument of the political movement hypothesis. For simplicity's sake, I present this as a linear relationship, though future research might loosen this assumption. It is based on the same 0–1 voting space. Rather than measure a country's absolute political alignment, however, it charts the change (or realignment) in that position from one time period to another. The maximum distance a country can move vis-à-vis the lender is ±1.0. Countries that make large movements toward the U.S., such as at point B, have a greater chance of receiving IMF credit than those that make movements away from the U.S., such as at point A. Figure 3 brings the spatial and temporal sides of the story together to illustrate the effect of political realignment from one voting cycle to the next. A country shifting from point A to point B has a better chance of receiving a loan than one moving from point D to point C, even though point C is still closer to the lender's position than is point B.

OPERATIONALIZATION OF THE HYPOTHESES

Because theory tells us that both economic and political factors affect IMF lending, a model that excludes either category is, by definition, misspecified. A combined political economy approach addresses one of Bird's main concerns:

> [S]ome countries may be able to muster more support amongst the membership of the Executive Board than others. The problem is that such [political] factors are difficult to model formally and include in econometric estimation, but their exclusion may explain why demand functions which rely exclusively upon economic characteristics will leave much of the story untold.[49]

It may be difficult to model and test political factors econometrically, but it is certainly possible.

[49] Bird (fn. 1, 1995), 149–50.

The analysis focuses not on the size of the loans, the interest rate, c other conditions but simply on the decision to lend. The structure c the Fund leaves more room for political factors to enter into the pro cess of loan approval than into the formation of the terms of the loan themselves. The politically appointed executive board generally vote yes or no on a complete package that has been assembled by the staff based largely upon market conditions and its own modeling and fore- casting. Furthermore, Fund rules on confidentiality make data collec- tion on most of the terms and conditions of loan packages impossible.[50]

This section presents the hypotheses introduced above in the forr of a pooled logit model of IMF lending to eighty-seven developing countries from 1985 to 1994.[51] The dichotomous nature of the depen- dent variable requires the use of logit estimation, which treats the rela- tionship between a categorical dependent variable (the probability of receiving a loan) and the continuous independent variables as a nonlin- ear one that approaches both 0 and 1 asymptotically.[52]

The basic logit can be expressed symbolically as:

$$\ln[(P(L = 1)_{it})/(1-(P(L = 1)_{it})] = b_0 - b_1(\text{BalPay}) + b_2(\text{Debt}) - b_3(\text{PerCapY}) - b_4(\text{Default}) + b_5(\text{USX}) + b_6(\text{USDFI}) + b_7(\text{PolProx}) + b_8(\text{PolMove})$$

L is a dichotomous variable coded 1 if country i received an IMF Stand-by Arrangement (SBA) or Extended Fund Facility (EFF) loan during the calendar year t, 0 if it did not.[53] $P(L = 1)_{it}$ is the estimated

[50] Regressions were run on the amount of the loan divided by GNP, and the general results were sim- ilar to those reported here, particularly for the political variables.

[51] This figure represents all of the developing countries, as defined by the IMF, for which data were available. See IMF, *Annual Report* (Washington, D.C.: IMF, 1986), 162. Data for the indicator of polit- ical alignment used here are not available before 1983. For some countries, data are available only for certain years. See Appendix B for a list of countries used in the data analysis.

[52] See John H. Aldrich and Forrest D. Nelson, *Linear Probability, Logit, and Probit Models* (Beverly Hills, Calif.: Sage, 1984).

[53] Two other IMF lending programs, the Structural Adjustment Facility (SAF) and the Enhanced Structural Adjustment Facility (ESAF), are not included in this analysis for a number of reasons. First, only low-income developing countries qualify for SAF and ESAF loans. A large number of countries in the sample would therefore not qualify for these programs, while all are eligible for SBA and EFF pack- ages. Second, the SAF and ESAF are structural adjustment rather than economic stabilization programs. To include them in the analysis would require a different underlying macroeconomic model than that specified for SABs and EFFs. Third, 1987 was the first full year of operation for the SAF and 1988 for the ESAF. Only SBA and EFF programs were operational throughout the entire time period examined here. See Polak (fn. 6, 1991); and Susan Schadler, Adam Bennett, Maria Carkovic, Louis Dicks-Mireaux, Maruo Mecagni, James H. J. Morisink, and Miguel A Savastano, "IMF Conditionality: Experience under Stand-By and Extended Arrangements. Part I: Key Issues and Findings," IMF Occasional Paper, no. 128 (1995). Compared to the number of SBAs and EEFs, there have been few SAF and ESAF loans

probability that a country will receive a loan in year *t*.[54] b_0 is the intercept term. "BalPay" is the balance of payments position; "Debt" is the debt burden; "PerCapY" is per capita income; "Default" is a measure of credit history; "USX" is the amount of U.S. exports to the country; "USDFI" is the amount of U.S. direct foreign investment in the country; "PolProx" is political proximity to the U.S. in the international voting space described above; and "PolMove" is political movement toward the U.S. within the same space. The economic variables are lagged by one year and the political variables by one to two years to establish the direction of causality.

To pose a more challenging test for the political model, I isolate a number of critical macroeconomic factors expected to affect Fund behavior and select several statistical proxies for them. The balance of payments is operationalized into several different variables. BOP_{it-1} is the overall balance of payments of a country in year *t-1*.[55] ΔBOP_{it} is the change in the overall balance of payments from *t-1* to *t*. $PCBOP_{it-1}$ and $\Delta PCBOP_{it}$ are per capita balance of payments and change in per capita balance of payments, respectively.[56] $CACCT_{it-1}$ is the current account, and $\Delta CACCT_{it}$ is the change in the current account. $CACCT/GNP_{it-1}$ and $\Delta CACCT/GNP_{it}$ capture the ratio of the current account to GNP and the change in that ratio.[57] Since higher payments deficits are thought to increase the chances of receiving a loan, all coefficients should be negative.

A country's debt burden is measured by the following variables: $DEBT_{it-1}$ is the level of absolute public and publicly guaranteed debt in year *t-1*, and $\Delta DEBT_{it}$ is the change in that level of debt from *t-1* to *t*. $PCDEBT_{it-1}$ and $\Delta PCDEBT_{it}$ compute per capita debt figures. A se-

made. Regressions run on a variable including all of these programs together yielded results generally consistent with those reported in the following section.

[54] Logit transforms this variable, which has a nonlinear relationship to the independent variables, into the log-odds of receiving a loan, which has a linear relationship to the independent variables. The new dependent variable, or logit, is then regressed on the independent variables using maximum likelihood estimation (MLE). Data for this variable were gathered from IMF, *Annual Report* (Washington, D.C.: IMF, various issues).

[55] All economic variables except ratios are expressed in millions of 1990 U.S. dollars, using the 1990 U.S. GDP deflator reported in IMF, *International Financial Statistics Yearbook*. (Washington, D.C.: IMF, various issues).

[56] These variables make the figures for large and small countries more comparable. I also tested the ratio of balance of payments to GNP and the change in that ratio with the same substantive results. Data are from IMF, *International Financial Statistics Yearbook* (Washington, D.C.: IMF, various issues).

[57] The World Bank's debt ratios (DEBT/GNP, INT/GNP, and RES/DEBT) appear to have been multiplied by 100. To make comparisons across units consistent, I multiplied the CACCT/GNP ratios calculated from (but not listed in) World Bank data by 100. World Bank, *World Debt Tables* (Washington, D.C.: World Bank, various issues); and idem, *Global Development Finance* (Washington, D.C.: World Bank, various issues).

ries of ratios captures the debt service burden for each econom $DEBT/GNP_{it-1}$ is the ratio of long-term total debt stocks (public ar private) to GNP, and $\Delta DEBT/GNP_{it}$ tracks the change in that rati INT/GNP_{it-1} and $\Delta INT/GNP_{it}$ are the variables for the ratio of inte est payments to GNP and its change, and $RES/DEBT_{it-1}$ an $\Delta RES/DEBT_{it}$ measure the ratio of reserves to GNP and the change i that ratio.[58] Because a heavy debt burden increases debtors' demand fi loans and because the Fund is hypothesized to give greater supply coı sideration to the larger debtors,[59] all coefficients should be positive e: cept those for $RES/DEBT_{it-1}$ and $\Delta RES/DEBT_{it}$, which are expecte to be negative.

$PCAPY_{it-1}$ represents per capita GNP, computed from data reporte in the World Bank's *World Debt Tables* and *Global Development Finan* and in the IMF's *International Financial Statistics*. Lower-income coun tries should be more likely to receive loans, so a negative coefficient expected.

$DEFAULT_{it-1}$ is a dummy variable coded 1 if a country has had prior IMF program canceled any time since 1975 (the first full year (the EFF program) through year *t-1*, 0 otherwise. Since a bad credit his tory should adversely affect the likelihood of future loans, its coefficie should be negative.[60]

USX_{it-1} is the level of U.S. exports to a country. According to th neomarxist hypothesis, it should be positively signed. The domesti level political economy perspective has ambiguous expectations for th direction of this effect.[61]

$USDFI_{it-1}$ is the value of the stock of U.S. direct investment for a industries in a country. Expectations are similar to those for USX_{it-1}.[6

The voting space depicted in Figures 1 to 3 is measured b $KVOTE_{it-2}$, an index of political agreement between country *i* and th U.S. in year *t-2*. Calculated as a decimal between 0 and 1, this variab measures the degree of coincidence between the votes of the samp

[58] These figures are from the World Bank, *World Debt Tables* (Washington, D.C.: World Bank, va ious issues); and idem, *Global Development Finance* (Washington, D.C.: World Bank, various issue with population data taken from IMF, *International Financial Statistics Yearbook* (Washington, D.C IMF, various issues).

[59] Adequate data on the exposure of U.S. banks in particular countries are unavailable. In any ever the largest creditor banks are likely to be based in the U.S. and the IMF's other principal sharehold countries.

[60] A variable measuring the total number of cancellations that a country experienced from 19 through *t–1* did not yield statistically significant results. Data were gathered from IMF, *Annual Rep* (Washington, D.C.: IMF, various issues).

[61] Data are from IMF, *Direction of Trade Statistics Yearbook* (Washington, D.C.: IMF, various issues

[62] Data have been taken from the U.S. Department of Commerce, *Survey of Current Business*, va ous issues.

country and the U.S. in the United Nations General Assembly (UNGA) on the approximately ten to fifteen issues in each session that the U.S. Department of State has deemed key votes. Under Congressional mandate, the State Department has compiled the voting records of all UN member nations on these selected issues since the 1983 General Assembly in its annual publication, *Report to Congress on Voting Practices in the United Nations*.[63] Appendix A lists the key votes identified in this report for the years examined here. In accordance with the political proximity hypothesis, a positive coefficient is anticipated.[64]

Using UN voting patterns to measure international political alignment is one solution to the problem of testing political variables lamented by Bird. For its part, the U.S. government has proclaimed that examining UN votes makes it "possible to make judgments about whose values and views are harmonious with our own, whose policies are consistently opposed to ours, and whose practices fall in between."[65] But not all UN votes are equally important. In reference to the key votes, the same report states that the "only votes that can legitimately be read as a measure of support for the United States are those which we identified as important to us, and on which we lobbied other nations"[66] The validity of UNGA voting records has been debated extensively.[67] I

[63] Using these annual reports, I coded votes in agreement with the U.S. 1.0, votes in disagreement with the U.S. 0.0, and abstentions or absences by the sample country 0.5. I then added and divided these numbers by the total number of key votes each year to come up with the annual decimal measure for each country. This method differs slightly from the technique of discarding absences and abstentions from the total count of UNGA votes used in Charles W. Kegley Jr. and Steven W. Hook, "U.S. Foreign Aid and U.N. Voting: Did Reagan's Linkage Strategy Buy Deference or Defiance?" *International Studies Quarterly* 35 (September 1991). Rather than not count those nonvotes on "key" issues, I interpret them as neutral.

[64] The transmission of United States foreign policy preferences from the State Department is not necessarily direct in the case of the multilateral development banks and the IMF, where Treasury plays a critical role. See Schoultz (fn. 10). The (American) deputy managing director has typically been "a Treasury man,' reinforcing the close ties between that agency of the U.S. government and the IMF." Kahler (fn. 16), 94. Furthermore, Kahler argues that Treasury maintains tight control over U.S.-Fund relations and that "other agencies that might attempt to politicize the IMF for broader foreign policy goals tended to be excluded from direct access to it." Kahler (fn. 16) 94, 97. On the other hand, Joanne Gowa notes that Treasury has adopted an ordering of priorities that "subordinates the demands of the international monetary order to the imperatives of domestic economic policy and foreign security policy," suggesting some coordination—or at least compatibility—between different agencies within the executive branch. Gowa, *Closing the Gold Window: Domestic Politics and the End of Bretton Woods* (Ithaca, N.Y.: Cornell University Press, 1983). The present analysis of policy output (as opposed to interagency input) is an indirect test of these two competing hypotheses. Future work should address the interagency dynamics more directly.

[65] U.S. Department of State, *Report to Congress on Voting Practices in the United Nations* (1985), 2.

[66] Ibid., 4.

[67] See Soo Yeon Kim and Bruce Russett, "The New Politics of Voting Alignments in the United Nations General Assembly," *International Organization* 50, no. 4 (1996); Steven K. Holloway and Rodney Tomlinson, "The New World Order and the General Assembly: Bloc Realignment at the UN in the Post–Cold War World," *Canadian Journal of Political Science* 28, no. 2 (1995); Leona Pallansch and

adopt the self-identified measure of political alignment of the principa U.S. foreign policy decision-making body: UNGA key votes. There is also evidence that the State Department tracked such data in a similar manner previous to the Congressional mandate, that it considered UN votes a reliable indicator of alignment, and that the U.S. allocated aid on the basis of that alignment. In a 1964 memo to the director of the Food for Peace Program, Dick Reuter, Lansdale noted that "at critical moments in the world's recent history, the U.S. 'bought' votes subtly and indirectly to support its stand in the General Assembly. The 'buying' is in terms of U.S. assistance to the voting country."[68] Furthermore, Lansdale's analysis employed a measure of alignment similar to the current State Department use of key votes, charting only votes on cold war issues.

It also appears that at least some recipient countries take U.S. vigilance of UN voting seriously. Argentina, for example, previously a leader in the Non-Aligned Movement, modified its voting stance in the UNGA to reflect better its improved relations with the United States in the early 1990s.[69] In a 1997 interview, Carlos Escudé, a former adviser to Argentina's minister of foreign relations, revealed that "with respect to some important United Nations resolutions, there was direct contact between Argentina and the United States, and Argentina voted in a manner favorable to the United States."[70] More generally, between 1990 and 1991 Argentina altered its UN votes to move from the fourth, most anti-U.S. stance in the UN to a position similar to that Turkey.[71]

MKVOTE$_{it-1}$ measures the movement in political alignment b tween the sample country and the U.S. within the voting space fro

Frank Zinni Jr., "Demise of Voting Blocs in the General Assembly of the UN? A Multidimensional Sc ing Analysis" (Paper presented at the annual meeting of the Southern Political Science Association, A lanta, 1996); Brian W. Tomlin, "Measurement Validation: Lessons from the Use and Misuse of U General Assembly Roll-Call Votes," *International Organization* 39, no. 1 (1985); and Kenneth Menkhaus and Charles W. Kegley Jr., "The Compliant Foreign Policy of the Dependent State Rev ited: Empirical Linkages and Lessons from the Case of Somalia," *Comparative Political Studies* 21, nc (1988).

[68] Ed Lansdale, "Memo Re: Long Range Impact FPF-II," April 24, 1964, National Archives, Recc Group 59, Lot file 67D554, Under Secretary for Political Affairs, Records of the Special Assists 1963–65, Box 2. I thank Tim McKeown for providing me with a transcription of this document.

[69] For example, Argentina sent troops to the 1991 Persian Gulf conflict. Carlos Escudé, "Entrevi a Escudé realizada por Lorena Kniaz" (1997), cited May 19, 1999, http://www.geocities.coı CapitolHill/Congress/4359/reporta.html.

[70] Ibid.

[71] Ibid.; Carlos Escudé, E-mail from the author, February 16, 1999.

rear *t-2* to year *t-1*, measured in UNGA key votes.[72] From the political novement hypothesis, I expect a positive coefficient.

FINDINGS AND INTERPRETATIONS

Table 1 presents the results of three different versions of the model. The combination of pooled data and a categorical dependent variable presents unique diagnostic challenges.[73] Column 1 presents the results or the basic logit model, with no correction for autocorrelation. It appears to provide a good overall fit: -2 times the log-likelihood ratio $-2 \times$ LLR) for the model is 124.85, with $p < 0.0001$. We can reject the null hypothesis that none of the independent variables individually or collectively explain a significant amount of variation of the dependent variable.[74] The model correctly predicted 83.25 percent of the outcomes. In terms of individual coefficients, $PCAPY_{it-1}$ and $KVOTE_{it-2}$ are significant at the 0.90 level of confidence; BOP_{it-1} and $PCDEBT_{it-1}$ at 0.95; and $DEBT/GNP_{it-1}$, INT/GNP_{it-1}, $\Delta INT/GNP_{it}$, $RES/DEBT_{it-1}$, and $MKVOTE_{it-1}$ at 0.99; all with the anticipated signs.

To test and correct for autocorrelation, I employed the binary time-eries-cross-section estimation technique formulated by Beck, Katz, nd Tucker.[75] I constructed a series of nine (T−1) dummy variables oded 1 if it had been (1, 2, 3, . . . T−1) years since a country last re-eived an IMF loan, 0 otherwise. If these nine variables collectively are gnificant in a log-likelihood ratio test, it is an indication of autocorre-tion. The correction for autocorrelation is simply the inclusion of the mporal dummy variables in the estimation. Once corrected, the new oefficients for the original variables of interest should better satisfy the

[72] Because of the UNGA's voting calendar, the voting variables have a longer lag structure than the onomic variables. The fact that UNGA votes are taken in the last four months of the calendar year eans that there is a 67 percent chance that a given loan decision will be made before the UNGA meets a given year. The chances that such a decision will be made before the session is complete and final tes are tallied approaches 100 percent. Conversely, movement at *t–1* occurs immediately before the xt calendar year's loan cycle begins.

[73] James A. Stimson, "Regression in Space and Time: A Statistical Essay," *American Journal of Polit-l Science* 29, no. 4 (1985); Nathaniel Beck, Jonathan N. Katz, and Richard Tucker, "Taking Time Se-usly: Time-Series-Cross-Section Analysis with a Binary Dependent Variable," *American Journal litical Science* 42, no. 4 (1998).

[74] This assumes a Chi Square distribution for the $-2 \times$ LLR figure. While this assumption may not entirely valid for individual level data, the strong results are still encouraging.

[75] Beck, Katz, and Tucker (fn. 73). This approach is designed for longitudinally dominant data with ically twenty or more time periods. The authors have not yet tested this exploratory method on orter time periods like the one used here ($T = 10$). Richard Tucker, conversation with the author, gust 1998. We may therefore have somewhat less confidence in a negative diagnostic for autocorre-ion than in the positive one obtained here.

TABLE 1

LOGIT COEFFICIENT ESTIMATES OF IMF LENDING, 1985–94

Independent Variables	*Basic Logit (1)*	*With Temporal Dummies (2)*	*Refined with Dummies (3)*
BOP_{it-1}	$-1.907 \times 10^{-4***}$	$-2.091 \times 10^{-4***}$	$-1.277 \times 10^{-4***}$
	(0.911×10^{-4})	(0.926×10^{-4})	(0.583×10^{-4})
ΔBOP_{it}	-1.200×10^{-6}	1.580×10^{-6}	
	(79.600×10^{-6})	(84.000×10^{-6})	
$PCBOP_{it-1}$	6.642×10^{-4}	8.886×10^{-4}	
	(17.944×10^{-4})	(18.001×10^{-4})	
$\Delta PCBOPt_{it}$	-7.526×10^{-4}	-9.787×10^{-4}	
	(18.621×10^{-4})	(19.023×10^{-4})	
$CACCT_{it-1}$	-1.089×10^{-4}	-1.252×10^{-4}	
	$(.956 \times 10^{-4})$	(0.974×10^{-4})	
$\Delta CACCT_{it}$	0.920×10^{-4}	1.192×10^{-4}	$1.453 \times 10^{-4***}$
	(0.976×10^{-4})	(1.031×10^{-4})	(0.632×10^{-4})
$CACCT/GNP_{it-1}$	–0.020	–0.022	–3.506*
	(0.026)	(0.027)	(2.044)
$\Delta CACCT/GNP_{it}$	0.002	0.003	
	(0.024)	(0.025)	
$DEBT_{it-1}$	-2.400×10^{-6}	-6.600×10^{-6}	
	(11.000×10^{-6})	(11.400×10^{-6})	
$\Delta DEBT_{it}$	-3.320×10^{-5}	-4.760×10^{-5}	
	(6.63×10^{-5})	(6.730×10^{-5})	
$PCDEBT_{it-1}$	$9.960 \times 10^{-4**}$	$8.726 \times 10^{-4**}$	$8.319 \times 10^{-4*}$
	(4.58×10^{-4})	(4.655×10^{-4})	(4.340×10^{-4})
$\Delta PCDEBT_{it}$	5.641×10^{-4}	1.525×10^{-4}	
	(12.84×10^{-4})	(13.253×10^{-4})	
$DEBT/GNP_{it-1}$	–0.011***	–0.010***	–0.010***
	(0.003)	(0.003)	(0.003)
$\Delta DEBT/GNP_{it}$	0.001	0.002	
	(0.003)	(0.004)	
INT/GNP_{it-1}	0.284***	0.274***	0.267***
	(0.065)	(0.069)	(0.065)
$\Delta INT/GNP_{it}$	0.503***	0.519***	0.516***
	(0.100)	(0.102)	(0.099)
$RES/DEBT_{it-1}$	–0.026***	–0.027***	–0.024**
	(0.010)	(0.010)	(0.010)
$\Delta RES/DEBT_{it}$	-4.541×10^{-4}	10.000×10^{-4}	
	(0.021)	(0.022)	
$PCAPY_{it-1}$	$-3.638 \times 10^{-4**}$	-3.074×10^{-4}	$-3.453 \times 10^{-4**}$
	(02.112×10^{-4})	(2.122×10^{-4})	(2.052×10^{-4})
$DEFAULT_{it-1}$	0.394*	0.412*	0.467**
	(0.231)	(0.238)	(0.226)
USX_{it-1}	-4.800×10^{-6}	-2.500×10^{-6}	
	(65.300×10^{-6})	(67.300×10^{-6})	

ABLE 1 *(cont.)*

dependent riables	*Basic Logit (1)*	*With Temporal Dummies (2)*	*Refined with Dummies (3)*
$SDFI_{it-1}$	-9.940×10^{-5}	-8.870×10^{-5}	
	(8.800×10^{-5})	(8.880×10^{-5})	
$VOTE_{it-2}$	1.247*	1.004	0.898
	(0.716)	(0.728)	(0.677)
$KVOTE_{it-1}$	2.756***	2.858***	2.711***
	(0.795)	(0.856)	(0.827)
tercept	−2.294***	−2.243***	−2.247***
	(0.476)	(0.522)	(0.490)
orrectly predicted (%)	83.25	82.98	82.98
odel χ^2 value	124.85 $p < 0.0001$ 24 d.f.	145.03 $p < 0.0001$ 33 d.f.	140.14 $p < 0.0001$ 21 d.f.

= 746. Standard errors are in parentheses below the estimates. Coefficients for temporal dummies ot reported.

* Significant at $p \leq 0.10$ level.

** Significant at $p \leq 0.05$ level.

*** Significant at $p \leq 0.01$

ssumption of error independence. This test revealed a high likelihood f autocorrelation (log-likelihood ratio = 20.18 with 9 d.f., $p < 0.025$). Columns 2 and 3 in Table 1 report the results of the full model and a nore refined model correcting for autocorrelation.

Tests were also conducted for multicollinearity and heteroskedasticy. Neither revealed any indications of problems. Despite the presence f a large number of potentially overlapping economic variables, none f the variables in the model exhibited high degrees of collinearity with he other variables (either collectively or individually). To test for hetroskedasticity, I incorporated a variable for GNP to test the impact of ountry size and a series of eight dummy variables to capture the effect f geographic region. The results of these tests did not approach conentional levels of statistical significance, so I retained the assumption f homoskedasticity.

The parameter estimates of the corrected full model (Column 2) ield several interesting, albeit tentative, findings. The only significant alance of payments variable is the overall balance, BOP_{it-1}. As exected, its coefficient is negatively signed, suggesting that a country

with an extra $100 million payments deficit increases its log-odds of receiving an IMF loan by about 0.02. The insignificant findings for the current account differ from others' results.[76] Controlling for the overall balance, the current account does not seem to matter. Similarly, with the current account controlled for, the negative impact of the balance of payments supports the argument that there is a substitution effect between IMF lending and other foreign capital inflows. Foreign capital inflows (an improvement in the balance of payments) in the year prior to the lending period lower the log-odds of receiving a loan from the IMF.

Several demand-side debt indicators were statistically significant.[77] In particular, the ratio of interest payments to GNP and the change in that ratio seem to have a strong positive impact on the log-odds of receiving a loan. The coefficients for the supply-side aggregate debt indicators $DEBT_{it-1}$ and $\Delta DEBT_{it}$ were not significant. These findings generally confirm the importance of debt in the borrowers' demand functions. They do not, however, support the argument that the IMF gives special treatment to large debtors, either because of their importance to global financial stability or as a payoff to the large creditor banks whose holdings may increase in value if an IMF loan is granted. Special treatment received by any particular debtors may be better explained by political factors than by their position in the international financial system or their relationship with creditor banks. This is a particularly interesting finding in light of the controversies surrounding the U.S.- and IMF-sponsored bailout packages in 1995 in Mexico and in 1997 and 1998 in Asia and Russia.[78]

Per capita income behaves as expected but is no longer significant in the corrected model ($p < 0.1475$). A country's history of default with the IMF is significant at the 0.90 level, but carries a positive sign. This contradicts the notion that economic neutrality drives Fund lending and confirms the pattern of recidivism observed by others.[79] Having had a previous IMF program canceled increases the log-odds of receiv-

[76] Conway (fn. 25).

[77] Per capita debt reached the 0.90 level of confidence, and the following variables attained the 0.95 level: the debt to GNP ratio, the interest to GNP ratio, the change in the interest to GNP ratio, and the ratio of reserves to debt. Curiously, the coefficient for debt to GNP is negative (all others are correctly signed). I have no explanation for this anomalous result, except to surmise that the impact of high relative levels of debt may be sensitive to the burden of repayment as captured by the interest to GNP ratio

[78] On the Mexican crisis, see Riordan Roett, "The Mexican Devaluation and the U.S. Response Potomac Politics, 1995-Style," in Roett, ed., *The Mexican Peso Crisis: International Perspectives* (Boulder, Colo.: Lynne Rienner, 1996).

[79] Bird (fn. 1, 1995).

g a new loan by 0.41.[80] The coefficients for U.S. exports and U.S. di-
ct foreign investment do not attain statistical significance at conven-
nal levels.[81] The neomarxist hypothesis is therefore not confirmed by
ese data. The potentially mixed interests of U.S. exporters and in-
stors described above makes the domestic politics argument more
fficult to assess. Some of the positive and negative impact of exports
d investments in different sectors of the economy (for example,
ded vs. nontraded, export vs. import competing, consumption vs. in-
rmediate goods) would be expected to work at cross purposes to yield
tistically insignificant results overall. Subtler model specification and
ture research may help clarify these issues.

Both political variables carry the correct sign, but the political prox-
ity hypothesis is not confirmed in the serially corrected model
< 0.1682). The results for the political movement hypothesis are
ongly positive and significant at the 0.99 level. A movement toward
e U.S. along the 0–1 UNGA key-vote continuum of 0.10 (for example,
itching one vote out of ten) raises the log-odds of receiving a loan by
29. Politics does matter but not in the manner typically argued. These
ta suggest that movement toward the U.S. within the political space
rtrayed in this paper influences IMF decisions regardless of absolute
gnment position. Additionally, the effects of these variables are ro-
st to changes in the specification of the underlying macroeconomic
odel. I do not report the intermediate results here, but the addition
d deletion of various economic variables had little effect on the para-
eter estimates or the standard errors of the political variables (see
ble 1, Columns 2 and 3).

I added an interactive political variable, $KVOTE_{it-2}*MKVOTE_{it-1}$, in
attempt to capture some of the potential nonlinear effects of re-
gnment by testing the hypothesis that the impact of a change in po-
ical alignment by a country is dependent on that country's starting
sition. A given movement toward the U.S. by an already tight Amer-
an ally, whose allegiance is unquestioned by the American govern-
ent, may not increase the probability of receiving a loan as much as
e same degree of movement by a more politically distant country. Ex-
ctations for the interactive term were tentative, but a negative coeffi-

[80] This result may be spurious. Bird suggests that requesting a loan from the IMF has a threshold ef-
t; once a country requests one loan, it is more likely to request additional loans. Since any country
t has a loan canceled has already crossed this threshold, it may be more likely to receive loans in the
ure. Bird (fn. 1, 1995).

[81] Regressions were also run using each variable without the other, yielding similar negative results.

cient would be consistent with this discussion. I do not report those results here, but the coefficient for this variable was not significant at conventional levels and positively signed.[82] The data suggest that the impact of movement toward the U.S. is consistent across different starting points. This supports the argument depicted in Figure 3 that realignment toward the U.S. improves a country's chances of receiving a loan regardless of the starting position.

Table 1, Column 3, presents the results of a refined statistical model, which largely confirm the above interpretations with the exception of the current account and per capita income variables. I constructed this model by sequentially deleting any previously nonsignificant variables and conducting a series of log-likelihood ratio tests to determine if their inclusion significantly improved the overall fit of the model. With the exception of $KVOTE_{it-2}$, which I retained because of its intrinsic interest, I omitted all variables not meeting these criteria from the refined model.[83] The overall current account balance still does not seem to matter, but its improvement or decline and its weight in the economy do. We detect some impact for the current account by eliminating some potentially overlapping variables.[84] Per capita income is negatively signed and significant at the 0.90 level, a modest improvement from the full model.

Because the logit model is nonlinear, the relative effect of any single variable depends on the value of all the independent variables, which determine where on the curve an estimate lies. To make the parameter estimates more readily interpretable, Table 2 uses the refined model results to illustrate the impact of different values of political realignment on the probability of receiving a loan from the IMF in the hypothetical case where the values of all other independent variables in the model are set at their means. Two clear patterns emerge. First, even if we assume that absolute alignment position matters, a political realignment has a much stronger impact on the probability of receiving an IMF lend-

[82] There was a possible collinearity problem with this variable. Specifically, it correlated strongly with $MKVOTE_{it-1}$. Because the inclusion of the interactive term is likely to have inflated the standard error of the movement variable and because its inclusion did not significantly improve the model's fit, I did not retain it.

[83] Because of the potential for omitted variable bias and the negative diagnostic for multicollinearity in the original specification, I have greater confidence in the results of the full model. I therefore present the refined model results for the interested reader but focus most of the substantive interpretations on the full model.

[84] The change in the current account from $t-1$ to t is significant at the 0.95 level, and the ratio of the current account to GNP is significant at the 0.90 level. The unexpectedly positive coefficient for the change in the current account from $t-1$ to t could be due to the fact that an IMF loan at time t can itself cause an improvement in the balance of payments at time t.

TABLE 2
THE EFFECT OF POLITICAL REALIGNMENT ON IMF LENDING[a]

	Original Position	*New Position*	*Loan Probability*
1.	0.0	0.0	0.065
2.	0.0	0.25	0.121
3.	0.0	0.50	0.213
4.	0.0	0.75	0.348
5.	0.0	1.0	0.513
6.	0.50	0.0	0.027
7.	0.50	0.25	0.053
8.	0.50	0.50	0.099
9.	0.50	0.75	0.178
10.	0.50	1.0	0.298
11.	1.0	0.0	0.011
12.	1.0	0.25	0.022
13.	1.0	0.50	0.042
14.	1.0	0.75	0.080
15.	1.0	1.0	0.147

Moments: $KVOTE_{it-1}$: mean = 0.5156, standard deviation = 0.1857;
$MKVOTE_{it-1}$: mean = 0.0082, standard deviation = 0.1555
[a]All other variables from Table 1, Column 3, held at their means.

ing package than the starting position.[85] Second, the patterns revealed in Table 2 are consistent with the scenario portrayed in Figure 3. A distant country that starts out at a key-vote index score of 0.0 and moves to 0.25 has a much better chance (p = 0.121) of receiving a loan than a country that moves away from perfect alignment with the U.S. (1.0) to a point (0.75) that is still much closer to the U.S. (p = 0.080) (Table 2, Lines 2 and 14). In fact, a country moving from discord to neutrality has a higher loan probability (p = 0.213) than a country that starts out and then remains in perfect alignment (p = 0.147) (Table 2, Lines 3 and 15).

Killick raises the possibility that changes in the structure of the international system alter the political dynamics treated here. Specifically, he suggests that the end of the cold war may dilute the effect of international politics on IMF behavior.[86] If no single power (for example, the USSR) lies at the other end of the 0–1 voting space, does the U.S. still reward political movement? *Does politics matter less after the cold war?*

[85] I retained absolute alignment position here to facilitate a clearer comparison of the hypothetical scenarios and to create more difficult conditions for demonstrating the strength of the impact of political realignment. Omitting $KVOTE_{it-1}$ would lower the probability for the static U.S. ally even more, relative to any country moving toward the U.S.
[86] Killick (fn. 1), 128.

TABLE 3

LOGIT COEFFICIENT ESTIMATES OF IMF LENDING, 1985–89 AND 1990–94

Independent Variables	*Cold War (1)*	*Post–Cold War (2)*	*Cold War Refined (3)*	*Post–Cold War Refined (4)*
BOP_{it-1}	$-4.072 \times 10^{-4**}$	-1.08×10^{-4}	$-4.331 \times 10^{-4**}$	-0.931×10^{-4}
	(1.736×10^{-4})	(1.623×10^{-4})	(1.72×10^{-4})	(1.596×10^{-4})
ΔBOP_{it}	-2.273×10^{-4}	-0.166×10^{-4}	$-2.784 \times 10^{-4**}$	0.304×10^{-4}
	(1.468×10^{-4})	(1.378×10^{-4})	(1.317×10^{-4})	(1.235×10^{-4})
$PCBOP_{it-1}$	0.004	3.88×10^{-4}	0.005^{*}	-2.057×10^{-4}
	(0.003)	(28.569×10^{-4})	(0.003)	(24.936×10^{-4})
$\Delta PCBOPt_{it}$	−0.002	0.002		
	(0.003)	(0.003)		
$CACCT_{it-1}$	-2.776×10^{-4}	-0.535×10^{-4}	-2.880×10^{-4}	-0.572×10^{-4}
	(1.855×10^{-4})	(1.529×10^{-4})	(1.777×10^{-4})	(1.523×10^{-4})
$\Delta CACCT_{it}$	3.040×10^{-4}	0.193×10^{-4}	3.023×10^{-4}	0.031×10^{-4}
	(1.857×10^{-4})	(1.311×10^{-4})	(1.844×10^{-4})	(1.274×10^{-4})
CACCT/	−0.046	−0.013	−0.054	−0.011
GNP_{it-1}	(0.039)	(0.044)	(0.036)	(0.044)
ΔCACCT/	−0.031	0.045	−0.042	0.053
GNP_{it}	(0.035)	(0.043)	(0.031)	(0.042)
$DEBT_{it-1}$	-1.93×10^{-5}	2.144×10^{-5}	-2.110×10^{-5}	1.976×10^{-5}
	(2.12×10^{-5})	(1.62×10^{-5})	(2.13×10^{-5})	(1.61×10^{-5})
$\Delta DEBT_{it}$	-1.639×10^{-4}	0.856×10^{-4}	$-1.591 \times 10^{-4*}$	0.875×10^{-4}
	(1.035×10^{-4})	(1.602×10^{-4})	(0.945×10^{-4})	(1.504×10^{-4})
$PCDEBT_{it-1}$	1.577×10^{-4}	0.002^{***}	1.205×10^{-4}	0.002^{***}
	(7.045×10^{-4})	(0.001)	(6.823×10^{-4})	(0.001)
$\Delta PCDEBT_{it}$	2.425×10^{-4}	0.001		
	(19.369×10^{-4})	(0.002)		
$DEBT/GNP_{it-1}$	−0.007	-0.016^{***}	−0.007	-0.016^{***}
	(0.005)	(0.005)	(0.005)	(0.005)
$\Delta DEBT/GNP_{it}$	-6.81×10^{-5}	0.001		
	(417.55×10^{-5})	(0.012)		
INT/GNP_{it-1}	0.362^{***}	0.183	0.363^{***}	0.197
	(0.094)	(0.142)	(0.092)	(0.135)
$\Delta INT/GNP_{it}$	0.530^{***}	0.646^{***}	0.530^{***}	0.655^{***}
	(0.153)	(0.184)	(0.151)	(0.173)
$RES/DEBT_{it-1}$	-0.056^{***}	−0.009	-0.052^{***}	−0.009
	(0.019)	(0.012)	(0.017)	(0.012)
$\Delta RES/DEBT_{it}$	−0.022	0.020		
	(0.044)	(0.028)		
$PCAPY_{it-1}$	-1.753×10^{-4}	$-7.171 \times 10^{-4**}$	-1.288×10^{-4}	$-7.021 \times 10^{-4**}$
	(3.345×10^{-4})	(3.188×10^{-4})	(3.096×10^{-4})	(3.16×10^{-4})
$DEFAULT_{it-1}$	0.447	0.345	0.438	0.350
	(0.322)	(0.404)	(0.320)	(0.400)

TABLE 3 *(cont.)*

Independent Variables	*Cold War (1)*	*Post–Cold War (2)*	*Cold War Refined (3)*	*Post–Cold War Refined (4)*
USX$_{it-1}$	1.757×10^{-4}	-1.041×10^{-4}	1.841×10^{-4}	-1.099×10^{-4}
	(1.141×10^{-4})	(1.508×10^{-4})	(1.131×10^{-4})	(1.528×10^{-4})
USDFI$_{it-1}$	-2.848×10^{-4}	-0.395×10^{-4}	-2.901×10^{-4}	-0.344×10^{-4}
	(1.972×10^{-4})	(1.036×10^{-4})	(1.997×10^{-4})	(1.02×10^{-4})
KVOTE$_{it-2}$	0.599	3.115**	0.566	2.967**
	(0.955)	(1.510)	(0.951)	(1.488)
MKVOTE$_{it-1}$	3.609**	4.333***	3.551**	4.192***
	(1.492)	(1.401)	(1.485)	(1.363)
Intercept	−1.949***	−3.570***	−2.008***	−3.468***
	(0.730)	(0.928)	(0.728)	(0.911)
Correctly predicted (%)	81.17	88.10	80.66	88.10
Model χ^2	96.75	72.04	95.81	70.64
p-value	$p < 0.0001$ 28 d.f.	$p < 0.0001$ 28 d.f.	$p < 0.0001$ 24 d.f.	$p < 0.0001$ 24 d.f.

N = 393 for cold war, N = 353 for post–cold war. Standard errors are in parentheses below the estimates. All specifications include temporal dummy variables (coefficients not reported).
*Significant at $p \leq 0.10$ level.
**Significant at $p \leq 0.05$ level.
***Significant at $p \leq 0.01$ level.

Table 3 presents the full and slightly refined results (using the same refining technique as above) of separate analyses for the 1985–89 and 1990–94 periods. The model provides a good fit for each of the two subsamples, with -2 × LLR for all four scenarios yielding $p < 0.0001$. The model had a higher success rate in predicting outcomes in the second period, correctly predicting 88 percent of the outcomes in both the full and restricted versions versus approximately 81 percent in both specifications for the first period.

The underlying macroeconomic models appear to differ slightly for the two periods. Balance of payments considerations have a greater impact during the cold war years (Table 3, Columns 1 and 3). The overall balance is correctly signed and significant at the 0.95 level. In the refined model, the change in the balance of payments is also significant at the 0.95 level and correctly signed. The coefficients for the current account and the change in the current account approach but do not quite attain statistical significance at the 0.90 level in both cold war specifications. No balance of payments variables even approach statistical sig-

nificance in the post–cold war period (Table 3, Columns 2 and 4). The relevant debt indicators for each subsample differ slightly, but both periods generally confirm the borrower need hypothesis with the exception of the incorrectly signed debt-to-GNP coefficient. The results for per capita income help clarify the ambiguous results for this variable in the full sample specifications by suggesting that while per capita GNP did not affect IMF decisions during the cold war, it has become more important in the post–cold war period. That confirms an apparent trend toward placing greater emphasis on economic growth in formulating IMF programs in recent years.[87] The default variable does not reach statistical significance in either period, possibly due to the smaller sample size. Again, the economic neutrality hypothesis is not confirmed. Finally, U.S. exports and U.S. direct foreign investment are not statistically significant in either subsample, though the USX_{it-1} variable comes reasonably close to attaining 90 percent confidence in the cold war period.

The differences between the impact of the economic variables across the two samples imply two tentative conclusions. First, similar models may produce divergent results if they are tested on different time periods. This may help explain the contradictory results of several seemingly similar econometric studies. Second, splitting larger time series into subsamples may be one good way to compare competing macroeconomic models and to chart their evolution over time.

The impact of politics also varies across the two subsamples but not in the way Killick anticipates. If anything, these results suggest that politics may be *more* important now than ever. The manner in which the U.S. treats its allies and potential allies within the Fund seems to have changed in important ways since 1990. The coefficient for alignment position does not approach statistical significance in the 1985–89 period, but movement is positively signed and statistically significant. Based on this sample, the U.S. appears to have been playing a cold war game of encouraging movement toward it without regard for original alignment position.

Since the end of the cold war, however, both alignment position and movement are statistically significant and positively signed. This suggests that the U.S. is both playing the realignment game as vigorously as ever and is rewarding the allegiance of those who stay close without necessarily moving any closer. Once a country reaches perfect agreement with the U.S., it cannot move any closer. These results imply that

[87] Killick (fn. 1).

during the cold war such a country would have had to move away from the U.S. and then back toward it to secure favorable treatment from the IMF. By rewarding such behavior, the United States may have encouraged countries to move toward the median voting position in the UN. Countries might also employ dual tactics of backscratching and blackmail to parlay political realignments and potential realignments into material gains.[88] Such maneuverings may no longer be necessary for close U.S. allies in the post–cold war period. More generally, these results suggest that the ability of the U.S. to influence IMF behavior to achieve its own political goals has not eroded. These goals may have simply shifted according to changes in the structure of the international system, and the U.S. still seems willing and able to exercise its weight within the executive board of the IMF to pursue them. The case of the IMF suggests that multilateralism, while useful for facilitating cooperation among a small number of like-minded states, may not be an effective buffer of U.S. power in the modern global political economy.

LIMITATIONS

This section highlights some of the limitations of this study's approach and data analysis with an eye toward future research. First and foremost, does voting in the UN General Assembly really matter, even on issues that the U.S. has deemed important? The UN itself has little power, and measures adopted within the UNGA in particular (as opposed to the Security Council) are largely symbolic.[89] In a similar vein, this study does not distinguish between countries according to their strategic and domestic characteristics. It could be argued that UN voting patterns are just a proxy for more fundamental variables. In particular, as countries democratize and open their economies to market forces, they may also be likely to alter their UN votes to reflect these underlying political and economic changes. The United States and IMF may be rewarding the political and economic shifts themselves, rather than the reflection of those shifts within the UNGA. This line of thinking is not necessarily inconsistent with the argument of this paper, but it merits further consideration. In fact, if UN voting does capture these more fundamental characteristics of countries, then it could be a very useful summary measure of them. I ran several new regressions to address these concerns empirically. I included commercial energy produc-

[88] See McKeown (fn. 48).

[89] Only one developing country (China) is a permanent member of the Security Council, so we cannot use Security Council votes to measure alignment.

tion ($ENERGY_{it-1}$) as a measure of strategic importance to explain why the U.S. might treat some countries differently from others. Measures of money supply (M/GDP_{it-1}), money supply growth ($MGROW_{it-1}$), budget deficits ($DEFICIT_{it-1}$), and trade openness ($OPEN_{it-1}$ = export plus imports, divided by GNP) captured the relative degree of "economic freedom."[90] Finally, several indicators of democracy, including the Polity III democracy (DEM_{it-1}) and authoritarianism ($AUTH_{it-1}$) scores and the Freedom House rankings on political rights (PR_{it-1}) and civil liberties (CL_{it-1}), helped assess the impact of regime type and democratization (change in regime type from one year to the next). Table 4 presents these results. Interestingly, none of these new variables yielded statistically significant results, and their inclusion in the estimation did not significantly alter the effects of the voting variables.[91] In sum, the model presented here appears robust to the inclusion of these factors.

Second, I have kept the underlying macroeconomic model as broad and simple as possible. This makes a direct comparison of theoretically distinct macroeconomic models more difficult, but the inclusion of a large number and wide range of economic variables raises the level of confidence in the statistical significance of the results obtained for the political variables, my more immediate concern. Further refinement or inclusion of additional economic variables could be undertaken if justified by other research.

Third, this paper treats the IMF essentially as an instrument of the U.S. government to test indirectly the proposition that relatively straightforward power considerations help explain the behavior of multilateral economic organizations. But the more powerful Fund members are likely to agree on many UNGA votes. Multidimensional scaling analyses conducted by Pallansch and Zinni suggest that the UNGA voting patterns of the G-7 countries tend to congregate together in a Euclidian space.[92] Future work should explore internal executive board politics and expand the focus to include Germany, Japan, France, and the United Kingdom. An approach centered around the formation and operation of subsets, or "k-groups," of countries within the organiza-

[90] Data from World Bank, *World Development Indicators 1998* (Washington, D.C.: World Bank, various issues); idem, *World Debt Tables* (Washington, D.C.: World Bank, various issues); and idem, *World Bank Global Development Finance* (Washington, D.C.; World Bank, various issues).

[91] The results for some of the economic variables differ from those in Table 1. Given the smaller number of cases used in Table 4 (a result of data availability), I base my substantive interpretations on the results presented in Tables 1 and 3. Several intermediate and refined specifications yielded similar results.

[92] Pallansch and Zinni (fn. 67).

tion could lead to a more complex specification of intraorganizational politics.[93]

Fourth, more careful consideration of the possible impact of U.S. domestic politics would help clarify and respecify those aspects of the problem. Specifically, the influence of domestic interest groups (for example, exporters and foreign investors) and the relations between different government agencies (particularly State and Treasury) merit further attention. Finally, I do not test directly for the impact of a country's past agreements with the IMF, nor do I exclude cases from the data sample that already have a program in effect. The former is partially captured by the default variable. The latter is much less of a problem than it appears because loans are often canceled and immediately replaced, suggesting that having a program in effect at a given moment does not exclude a country from the eligible sample.

IMPLICATIONS

This paper has two central goals: 1) to determine the degree to which high politics affects IMF lending patterns; and 2) to develop and test a more precise and more general explanation of *how* high politics influences the behavior of multilateral organizations. Most researchers of the politics of IMF lending argue that the U.S. punishes enemies and rewards friends via its influence within the Fund's executive board. Those who introduce a somewhat greater degree of complexity do not adequately develop nor test the dynamic impact of international political realignment. Previous research on foreign aid more generally has attempted but generally failed to find a statistically significant relationship between aid flows and political conditionality.[94] This paper provides the first systematic evidence that politics does affect IMF lending, and its conceptual framework and statistical analysis demonstrate the political factors that are most important, the mechanisms through which they influence Fund behavior, and the more general relationship between multilateral organizations and their member states. The results obtained here show that movement toward the United States within a defined international political space (like that measured by UN voting patterns) can significantly increase a country's chances of receiving a loan from the IMF. This suggests that the U.S. has been more concerned with attracting new allies and punishing defectors than reward-

[93] Ruggie (fn. 2).

[94] See McKeown (fn. 48).

Table 4
Logit Coefficient Estimates with Control Variables, 1985–94

Independent Variables	*Using Polity III Measures* *(1)*	*Using Freedom House Measures* *(2)*
BOP_{it-1}	-1.201×10^{-4}	-1.041×10^{-4}
	(1.184×10^{-4})	(1.192×10^{-4})
ΔBOP_{it}	6.235×10^{-5}	7.452×10^{-5}
	(11.02×10^{-5})	(11.19×10^{-5})
$PCBOP_{it-1}$	–0.002	–0.002
	(0.003)	(0.003)
$\Delta PCBOP_{it}$	–0.003	–0.003
	(0.003)	(0.003)
$CACCT_{it-1}$	9.496×10^{-5}	9.373×10^{-5}
	(13.09×10^{-5})	(13.14×10^{-5})
$\Delta CACCT_{it}$	2.163×10^{-4}	2.052×10^{-4}
	(1.332×10^{-4})	(11.327×10^{-4})
$CACCT/GNP_{it-1}$	–0.082*	–0.089*
	(0.049)	(0.049)
$\Delta CACCT/GNP_{it}$	–0.020	–0.021
	(0.043)	(0.043)
$DEBT_{it-1}$	1.933×10^{-5}	2.475×10^{-5}
	(2.01×10^{-5})	(2.00×10^{-5})
$\Delta DEBT_{it}$	-3.45×10^{-5}	-2.97×10^{-5}
	(8.35×10^{-5})	(8.21×10^{-5})
$PCDEBT_{it-1}$	0.001	0.002*
	(0.001)	(0.001)
$\Delta PCDEBT_{it}$	0.001	0.001
	(0.002)	(0.002)
$DEBT/GNP_{it-1}$	–0.014**	–0.016**
	(0.005)	(0.005)
$\Delta DEBT/GNP_{it}$	9.460×10^{-4}	6.491×10^{-4}
	(62.071×10^{-4})	(61.295×10^{-4})
INT/GNP_{it-1}	0.168	0.153
	(0.107)	(0.106)
$\Delta INT/GNP_{it}$	0.642***	0.663***
	(0.160)	(0.160)
$RES/DEBT_{it-1}$	–0.030*	–0.026*
	(0.018)	(0.017)
$\Delta RES/DEBT_{it}$	–0.026	–0.022
	(0.035)	(0.035)
$PCAPY_{it-1}$	-4.716×10^{-4}*	-6.376×10^{-4}*
	(4.072×10^{-4})	(3.833×10^{-4})
$DEFAULT_{it-1}$	0.6[illegible]3*	0.684*
	(0.361)	(0.364)
USX_{it-1}	1.079×10^{-4}	1.09×10^{-4}
	(0.928×10^{-4})	(0.897×10^{-4})

TABLE 4 *(cont.)*

Independent Variables	*Using Polity III Measures* (1)	*Using Freedom House Measures* (2)
$USDFI_{it-1}$	-1.328×10^{-4}	-1.452×10^{-4}
	(1.03×10^{-4})	(1.037×10^{-4})
$KVOTE_{it-1}$	1.095	1.350
	(1.209)	(1.203)
$MKVOTE_{it-1}$	4.138^{***}	4.464^{***}
	(1.270)	(1.265)
$ENERGY_{it-1}$	-1.11×10^{-5}	-1.13×10^{-5}
	(0.86×10^{-5})	(0.85×10^{-5})
M/GDP_{it-1}	0.017	0.015
	(0.012)	(0.011)
$MGROW_{it-1}$	-1.032×10^{-4}	-0.715×10^{-4}
	(3.026×10^{-4})	(3.033×10^{-4})
$DEFICIT_{it-1}$	0.062	0.062
	(0.050)	(0.050)
$OPEN_{it-1}$	-0.538	-0.490
	(0.694)	(0.688)
DEM_{it-1}	-0.033	
	(0.103)	
$AUTH_{it-1}$	-0.055	
	(0.123)	
PR_{it-1}		0.099
		(0.168)
CL_{it-1}		-0.134
		(0.214)
Intercept	-2.253^{*}	-2.533^{**}
	(1.247)	(1.170)
Correctly predicted (%)	83.72	84.84
Model χ^2	99.94	106.79
p-value	$p < 0.0001$	$p < 0.0001$
	40 d.f.	40 d.f.

N = 436 for Column 1, N = 455 for Column 2. Standard errors are in parentheses below the estimates. Both specifications include temporal dummy variables (coefficients not reported).

*Significant at $p \leq 0.10$ level.

**Significant at $p \leq 0.05$ level.

***Significant at $p \leq 0.01$ level.

ing loyal friends. It has been able to do so through multilateral channels like the IMF.

The evidence presented here also suggests that changes in the structure of the international system may have altered U.S. and IMF behavior but not in the predicted manner. In fact, these initial results suggest

that the end of the cold war has been associated with the increasing politicization of the IMF by the U.S. There is evidence that the U.S. has been willing to reward friends and punish enemies only since 1990. During the cold war (at least in its last few years), unless they were moving closer to the U.S. politically, allies of the U.S. had no greater chance than its adversaries of receiving assistance from the Fund. Only in the post-cold war period have these countries been able to cash in on their political allegiance.

The demonstration of the systematic impact of international politics on IMF lending poses interesting methodological, theoretical, and practical implications. Methodologically, the use of key UNGA votes provides a more easily quantifiable and temporally sensitive alternative to traditional indicators of international political alignment, such as security alliances, military base locations, treaties, and content analysis. The use of this indicator may facilitate research in other areas of inquiry.

On a theoretical level, the evidence presented here suggests that multilateral organizations like the IMF, despite their enhanced influence in the developing world, are still most profitably analyzed within the parameters of an international political context shaped primarily by the industrialized nations. More specifically, there is strong evidence that the political interests of the United States drive much of the behavior of one of the most important multilateral organizations in the post-hegemonic global economy. I do not explore the reverse causal relationship—the impact of the IMF on U.S. interests and behavior[95]—but these results suggest more generally that the multilateral institutions are still quite sensitive to direct political pressures and influences from their more powerful member states. These influences translate into particular modes of behavior by the multilateral organizations themselves that can be analyzed conceptually, observed empirically, and tested statistically. The study of the role of international institutions and multilateral organizations must take into account not simply the fact that international political factors help determine their behavior on the input side. Such research should also view the operation of such entities as a tool used by the great powers to achieve specific, identifiable, political goals on the output side, such as realignment within the international system. The ability of the U.S. to employ such tools underscores the practical limits of multilateralism and confirms the rather dramatized fears of one of the original architects of the postwar international economic order, John Maynard Keynes:

[95] See Kahler (fn. 16), 93

> There is scarcely any enduringly successful experience of an international body which has fulfilled the hopes of its progenitors. Either an institution has become diverted to the instrument of a limited group or else it has been a puppet — sawdust through which the breath of life does not blow.[96]

On the practical side, the experience of the IMF suggests that Keynes's first fear has been partially realized. To an extent, the U.S. has been able to use the IMF to further its own international political agenda. On perhaps a more positive note, his second fear of irrelevance appears to be a distant one, despite the relative economic decline of the U.S. and the end of the cold war. While undermining the principle of multilateralism, the continued strength of national influence over Fund behavior may well help maintain the stability of great power support for the multilateral organizations if those powers continue to reap important gains from them that may be more economically or politically costly to obtain bilaterally.[97] Such conclusions could help allay the fears of those within the U.S. Congress who question U.S. support for the IMF based on concerns that it would strengthen multilateralism at the expense of U.S. power.

Finally, as a multilateral organization, the IMF is in a sense a difficult or crucial case for political theories of international finance. It is easy to see how bilateral capital flows could be subjected to the push and pull of international and domestic politics, but on the executive board of the IMF any single country's power is diluted by the presence of other principals within the decision-making body. The structure of the Fund leaves the door open, but a priori we would expect to see less of an impact for politics in the IMF than in bilateral financial flows. If high politics affects IMF lending, then it should have an even stronger impact on national policies. A confirmation of the impact of political realignment on IMF lending therefore provides stronger corroboration of this theory than that which could be obtained in a study of bilateral capital flows and suggests that such ideas may be fruitfully applied to other areas of international finance and international relations more generally.

[96] Cited in Nick Butler, *The IMF, Time for Reform* (London: Fabian Society, 1982), 24.
[97] Cf. Ruggie (fn. 2), chap. 1.

Appendix A

UNGA Key Votes as Identified by the Department of State, 1983–93

Year	*Issue*	*Resolution*	*Vote (Yes-No-Abstain-Absent)*[a]
1983	Israeli credentials	Motion	79(U.S.)-43-19
	Middle East situation	38/180E	81-27(U.S.)-17
	Afghanistan	38/29	116(U.S.)-20-17
	Kampuchea	38/3	105(U.S.)-23-19
	Grenada	Motion	60-54(U.S.)-24
	Grenada	38/7	108-9(U.S.)-27
	Chemical and bacteriological weapons	38/187C	97(U.S.)-20-30
	Human rights in El Salvador	38/101	84-14(U.S.)-45
	Collaboration with South Africa	38/39G	122-9(U.S.)-17
	South Africa	38/39A	124-16(U.S.)-10
1984	Kampuchea	39/5	110(U.S.)-22-18
	Afghanistan	39/13	119(U.S.)-20-18
	Israeli credentials	Motion	80(U.S.)-41-22
	Chemical and bacteriological weapons	39/65A	99(U.S.)-14-13
	Military activities in dependent areas	39/412	62(U.S.)-47-24
	Apartheid	Motion	50-56(U.S.)-28
	Middle East	39/146	28(U.S.)-69-23
	Human rights in El Salvador	39/119	93-11(U.S.)-40
	Economic commission for Africa Conference Center	39/236 (III)	83-3(U.S.)-13
	Middle East situation	39/146A	69-39(U.S.)-26
1985	Kampuchea	40/7	114(U.S.)-21-16
	Afghanistan	40/12	122(U.S.)-19-12
	Human rights in Afghanistan	40/137	80(U.S.)-22-40
	Human rights in Iran	40/141	53(U.S.)-30-45
	Israeli credentials	Motion	80(U.S.)-41-20
	Chemical and bacteriological weapons	40/92C	112(U.S.)-16-22
	Namibia	Vote to retain 40/97B	54-63(U.S.)-29
	Middle East	Vote to retain 40/168A	64-33(U.S.)-41
	Central America	40/188	91-6(U.S.)-49
	Budget	40-253A	127-10(U.S.)-11
1986	Kampuchea	41/6	115(U.S.)-21-13
	Israeli credentials	Motion	77(U.S.)-40-16
	Nicaragua	41/31	94-3(U.S.)-47
	Afghanistan	41/33	122(U.S.)-20-11
	Libya	41/38	79-28(U.S.)-33
	Namibia	Vote to retain 41/39A	57-46(U.S.)-40
	Chemical and bacteriological weapons	41/58C	137(U.S.)-0-14
	Supplemental budget	41/211	122-13(U.S.)-10
	Human rights in Afghanistan	41/158	89(U.S.)-24-36

Year	Issue	Resolution	Vote (Yes-No-Abstain-Absent)[a]
	Middle East	Vote to retain 41/162A	66-38(U.S.)-41
1987	Israeli credentials	Motion	80(U.S.)-39-10-30
	Kampuchea	42/3	117(U.S.)-21-16-5
	Trade embargo against Nicaragua	42/176	94-2(U.S.)-48-15
	Afghanistan	42/15	123(U.S.)-19-11-6
	Human rights in Iran	42/136	64(U.S.)-22-45-28
	Apartheid	Vote to retain 42/23C	78-38(U.S.)-27-16
	Comprehensive system of international peace and security	42/93	76-12(U.S.)-63-8
	Program budget for the Biennium 1988–89	42/226	146-1-3(U.S.)-9
	Human rights in Afghanistan	42/135	94(U.S.)-22-31-12
	Middle East	Vote to retain 42/209B	64-33(U.S.)-41-21
1988	Israeli credentials	Motion	95(U.S.)-41-7-16
	Comply with International Court of Justice verdict in Nicaragua vs. U.S.	43/11	89-2(U.S.)-48-20
	Condemn foreign intervention in Cambodia	43/19	122(U.S.)-19-13-5
	Critical of human rights abuses in Iran	43/137	61(U.S.)-25-44-29
	Change name of PLO to "Palestine" in UN usage	43/177	104-2(U.S.)-36-17
	USSR resolution on international peace and security	43/89	97-3(U.S.)-45-14
	External debt crisis and development	43/198	150-1(U.S)-1-7
	Foreign intervention in Afghanistan	43/20	Adopted by consensus[b]
	Genuine and periodic elections	43/157	Adopted by consensus[b]
	Program budget outline	43/214	Adopted by consensus[b]
1989	Israeli credentials	Motion	95(U.S.)-37-15
	Situation in Kampuchea	44/22	124(U.S.)-17-12
	Situation in the Middle East	Paragraph vote	63-35(U.S.)-47
	Situation in the Middle East: Palestine and International Peace Conference	44/40A	109-18(U.S.)-31
	Situation in the Middle East: Golan Heights	44/40B	84-22(U.S.)-49
	International Court of Justice judgment re: Nicaragua	44/43	91-2(U.S.)-41
	UNRWA: Assistance to Palestine refugees	44/47A	134(U.S.)-0-1

Year	*Issue*	*Resolution*	*Vote (Yes-No-Abstain-Absent)*[a]
	Questions relating to information	44/50	127-2(U.S.)-21
	Cessation of all nuclear test explosions	44/105	136-3(U.S.)-13
	Amendment of the Limited Test Ban Treaty	44/106	127-2(U.S.)-22
	Prevention of an arms race in outer space	44/112	153-1(U.S.)-0
	Nuclear arms freeze	44/117D	136-13(U.S.)-5
	Indian Ocean Zone of Peace	44/120	137-4(U.S.)-14
	Enlargement of the Commission on Human Rights	44/167	151-2(U.S.)-2
	Trade embargo against Nicaragua	44/217	82-2(U.S.)-47
	Military intervention in Panama	44/240	75-20(U.S.)-40
1990	Comprehensive Nuclear Test Ban Treaty	45/51	140-2(U.S.)-6
	Bilateral nuclear arms negotiations	45/58H	99(U.S.)-0-50
	UNRWA: Assistance to Palestine refugees	45/73A	146(U.S.)-0-1
	Situation in the Middle East	Paragraph vote	52-37(U.S.)-49
	Situation in the Middle East: Palestine and International Peace Conference	45/83A	99-19(U.S.)-32
	Situation in the Middle East: Golan Heights	45/83B	84-23(U.S.)-41
	Periodic and genuine elections—UN electoral assistance	45/150	129(U.S.)-8-9
	Human rights in Occupied Kuwait	45/170	144(U.S.)-1-0
	Entrepreneurship	45/188	138(U.S.)-1-0
1991	IAEA report	Motion	88(U.S.)-25-26
	IAEA report	46/16	141(U.S.)-0-9
	Comprehensive Nuclear Test Ban Treaty	46/29	147-2(U.S.)-4
	Register of conventional arms transfers	46/36L	150(U.S.)-0-2
	UNRWA	46/46A	137(U.S.)-0-1
	Palestine—International Peace Conference	46/75	104-2(U.S.)-43
	Middle East—Palestinian question	46/82A	93-27(U.S.)-37
	Zionism/racism	Motion	34-96(U.S.)-13
	Zionism/racism	46/86	111(U.S.)-25-13
	Human rights in Occupied Kuwait	46/135	155(U.S.)-1-0
	Periodic and genuine elections	46/137	134(U.S.)-4-13
	Political and economic coercion	46/210	97-30(U.S.)-9
1992	Yugoslavia: UN membership	47/1	127(U.S.)-6-26
	IAEA report	47/8	146(U.S.)-0-5
	U.S. embargo of Cuba	47/19	59-3(U.S.)-71
	Comprehensive Nuclear Test Ban Treaty	47/47	159-1(U.S.)-4
	Maintenance of international security	47/60B	79(U.S.)-0-84
	Middle East—Golan Heights	47/63A	72-3(U.S.)-70
	Palestine—International Peace Conference	47/64D	93-4(U.S.)-60

ır	Issue	Resolution	Vote (Yes-No-Abstain-Absent)[a]
	UNRWA	47/69A	136(U.S.)-0-2
	Israeli practices	47/70A	83-5(U.S.)-55
	Situation in Bosnia and Herzegovina	47/121	102(U.S.)-0-57
	Periodic and genuine elections	47/138	141(U.S.)-0-20
	Human rights in Cuba	47/139	69(U.S.)-18-64
	Human rights in Sudan	47/142	104(U.S.)-8-33
	Human rights in Iraq	47/145	126(U.S.)-2-26
	Human rights in Iran	47/146	86(U.S.)-16-38
	External debt problems and development	47/198	158-1(U.S.)-0
›3	IAEA report	48/14	140(U.S.)-1-9
	U.S. embargo of Cuba	48/16	88-4(U.S.)-57
	Middle East peace process	48/58	155(U.S.)-3-1
	Middle East—Golan Heights	48/59B	65-2(U.S.)-83
	Israeli nuclear armament	48/78	53-45(U.S.)-65
	Situation in Bosnia and Herzegovina	48/88	109(U.S.)-0-57
	Periodic and genuine elections	48/131	153(U.S.)-0-13
	Human rights in Cuba	48/142	74(U.S.)-20-61
	Human rights in Iraq	48/144	116(U.S.)-2-43
	Human rights in Iran	48/145	74(U.S.)-23-51
	Human rights in Sudan	48/147	111(U.S.)-13-30
	Peaceful settlement of Palestine question	48/158D	92-5(U.S.)-51
	External debt problems of developing countries	48/182	164-1(U.S.)-0

›URCE: U.S. Department of State, *Report to Congress on Voting Practices in the United Nations*, vari-s years.

[a] U.S. vote shown in parentheses. Abstentions and absences combined from 1983–86. Only yes--abstain votes given for 1989–93.

[b] No records kept of which member states were present for consensus resolutions. Not included in :ing index scores of countries.

Appendix B
Countries Used in the Analysis

geria, Argentina, Bahamas, Bangladesh, Barbados, Benin, Bolivia, Botswana, Brazil, ırma/Myanmar, Burundi, Cameroon, Central African Rep., Chad, Chile, China, ›lombia, Comoros, Congo, Costa Rica, Côte d'Ivoire, Cyprus, Dominican Rebublic, uador, Egypt, El Salvador, Equatorial Guinea, Ethiopia, Fiji, Gabon, Gambia, Ghana, ·enada, Guatemala, Guyana, Haiti, Honduras, India, Indonesia, Jamaica, Jordan, nya, Lesotho, Liberia, Madagascar, Malawi, Malaysia, Mali, Malta, Mauritania, auritius, Mexico, Morocco, Nepal, Nicaragua, Niger, Nigeria, Oman, Pakistan, nama, Papua New Guinea, Paraguay, Peru, Philippines, Romania, Rwanda, Senegal, :rra Leone, Somalia, Sri Lanka, Sudan, Swaziland, Syria, Tanzania, Thailand, Togo, ınidad and Tobago, Tunisia, Turkey, Uganda, Uruguay, Venezuela, Yemen Arab public, P.D. Republic of Yemen, Yugoslavia, Zaire, Zambia

DEMOCRACY, INDIGENOUS *MOVEMENTS*, AND THE POSTLIBERAL CHALLENGE IN LATIN AMERICA

By DEBORAH J. YASHAR*

RECENT scholarship on third-wave democracies has come to focus on consolidation. After a decade of debates about the uncertainty of democratic transitions, scholars have become much less tentative about the prospects for democracy. Rather than study the sociopolitical processes by which social forces and states shape, support, and/or jeopardize the terms and direction of democracy, scholars have returned to an older intellectual tradition of comparing institutional differences between different governments and party systems. Accordingly, they explain the capacity to consolidate democracy largely as a function of institutional design.

This article takes issue with the conceptual and analytical underpinnings of the democratic consolidation literature. Specifically, it highlights how new political institutions, rather than securing democratic consolidation across the board, have in fact had a more checkered effect—as evidenced by the incomplete reach of the state, the survival of

* An earlier version of this paper was presented at Harvard University's Sawyer Seminar and Seminar on Ethics and International Affairs; at Cornell University's Sawyer Seminar; and at the 1998 Latin American Studies Association. I would particularly like to thank Oliver Avens, Eva Bellin, Sheri Berman, Jorge I. Domínguez, Kent Eaton, John Gershman, Jeffrey Herbst, José Antonio Lucero, Atul Kohli, Tali Mendelberg, María Victoria Murillo, Anna Seleny, Kathryn Stoner-Weiss, Sidney Tarrow, Donna Lee Van Cott, and the anonymous reviewers for their insightful comments.

Research conducted in Ecuador (November 1995 and February–May 1997), Bolivia (October 1995 and May–August 1997), Mexico (July–August 1996), Guatemala (March 1989–February 1990, December 1992, and February 1996), and Peru (August 1997) was supported by the Joint Committee on Latin American Studies of the Social Science Research Council and the American Council of Learned Societies; the United States Institute for Peace; the Helen Kellogg Center for International Studies at the University of Notre Dame; and at Harvard University, the Weatherhead Center for International Affairs, David Rockefeller Center for Latin American Studies, and the Milton Fund.

[1] This argument draws on Joel S. Migdal, *Strong Societies and Weak States: State-Society Relations and State Capabilities in the Third World* (Princeton: Princeton University Press, 1988); Joel Migdal, Atul Kohli, and Vivienne Shue, eds., *State Power and Social Forces: Domination and Transformation in the Third World* (Cambridge: Cambridge University Press, 1994); and Guillermo O'Donnell, "On the State, Democratization and Some Conceptual Problems: A Latin American View with Glances at Some Postcommunist Countries," *World Development* 21 (August 1993).

World Politics 52 (October 1999), 76–104

authoritarian enclaves, the uneven incorporation of social sectors, and the emergence of opposing social forces.[1] Indeed, many ostensibly consolidated democracies now find themselves being challenged by movements rallying against the failure of states to universalize democratic practices and secure political autonomy. Ethnic movements, in particular, have come increasingly to contest the foundations and contours of contemporary democratic and liberal institutions. These emergent movements have sparked fundamental political debates over territorial autonomy, legal pluralism, citizenship, representation, and multiculturalism.

These developments are particularly striking and consequential in Latin America. The regions' third-wave democracies have experienced increasing politicization of indigenous identities and organization of indigenous movements,[2] phenomena that appear to reverse the region's comparative historical record of weakly politicized ethnic cleavages.[3] These movements are most prominent in countries with large and moderate-size indigenous populations (Bolivia, Guatemala, Ecuador, and Mexico), but they have also provoked important debates and reforms in countries with small indigenous populations (Colombia, Brazil, and Chile). Indigenous movements have engaged in mass mobilizations, roadblocks, electoral campaigns, and policy negotiations. Unlike the class-based guerrilla wars of decades past, however, indigenous activists and movements do not seek to overthrow the state but rather are looking to reform democracy.

This article analyzes the politicization of ethnic cleavages as a springboard for delineating why the theoretical expectations of the democratic consolidation literature have not been realized; it then proposes how one might better theorize about democratic politics. It focuses on Latin America, the least likely case, where rural ethnic movements have come to assume increasing importance in the new de-

[2] See Rodolfo Stavenhagen, "Challenging the Nation-State in Latin America," *Journal of International Affairs* 45 (Winter 1992); Richard Chase Smith, "A Search for Unity within Diversity: Peasant Unions, Ethnic Federations, and Indianist Movements in the Andean Republics," in Theodore MacDonald Jr., ed., *Native Peoples and Economic Development* (Cambridge: Cultural Survival, 1985); Donna Lee Van Cott, ed., *Indigenous Peoples and Democracy in Latin America* (New York: St. Martin's Press, 1994); Deborah J. Yashar, "Indigenous Protest and Democracy in Latin America," in Jorge I. Domínguez and Abraham F. Lowenthal, eds., *Constructing Democratic Governance: Latin America and the Caribbean in the 1990s—Themes and Issues* (Baltimore: Johns Hopkins University Press, 1996); idem, "Contesting Citizenship: Indigenous Movements and Democracy in Latin America," *Comparative Politics* 31 (October 1998).

[3] See Crawford Young, *The Politics of Cultural Pluralism* (Madison: University of Wisconsin Press, 1976), chap. 11; Donald L. Horowitz, *Ethnic Groups in Conflict* (Berkeley: University of California Press, 1985); Ted Robert Gurr with Barbara Harff, Monty G. Marshall, and James R. Scarritt, *Minorities at Risk: A Global View of Ethnopolitical Conflicts* (Washington, D.C.: United States Institute of Peace Press, 1993).

mocracies, despite a contemporary history of weakly politicized ethnic cleavages.[4] The first section explains the increasing political salience of indigenous identities in twentieth-century Latin America by analyzing the unintended consequences of changing state-society relations. I argue that neoliberal-inspired citizenship reforms throughout the region have unintentionally challenged local autonomy, politicized ethnic identity, and catalyzed indigenous movements. The second section discusses the "postliberal challenge" that new indigenous movements pose for Latin America's third-wave democracies and for liberal state formation. Insofar as these movements demand new forms of representation, political autonomy, and multicultural recognition, they have once again engaged Latin America in a struggle over the kinds of democracies that will be built; the rights, responsibilities, and identities of citizens; and the ties that bind citizens to the state. Democratic institutions are therefore anything but consolidated. Rather, they are the subject of fundamental debates and the locus of far-reaching reforms, as evidenced by political negotiations and legislation throughout the region.

The final section discusses the implications of the prior analysis for how one studies politics in the region's new democracies. It questions the literature's conceptual focus on democratic consolidation and its analytic assumption of the power of new governmental institutions to define new interests, identities, and behavior uniformly and single-handedly. If conceptually this literature misconstrues regime endurance for consolidation, analytically it privileges governmental institutions as independent variables without explicitly evaluating alternative factors. As such it mistakenly *assumes* not only a competent and capable state but also a preconstituted society that will respond

[4] This article restricts its focus to indigenous identities, communities, and movements in the countryside, where the majority of self-identified Indians continue to reside. Most communities and movements have developed urban and occasionally international ties through markets, migration, and the rise of participation by nongovernmental organizations. However, even in these cases of increasing urban and international penetration, the *communities* in question remain geographically *rural*. This article does *not* address the politicization of racial cleavages in Latin America. The politicization of black identities, which has emerged in a different historical context, has been largely limited to urban movements and has resulted in types of political demands that are different from those voiced by indigenous movements in Latin America. For competing views on racial cleavages and their politicization, see Carl N. Degler, *Neither Black nor White: Slavery and Race Relations in Brazil and the United States* (Madison: University of Wisconsin Press, 1971); Pierre-Michel Fontaine, ed., *Race, Class, and Power in Brazil* (Los Angeles: University of California, 1985); Howard Winant, "Rethinking Race in Brazil," *Journal of Latin American Studies* 24 (February 1992), 173–92; Richard Graham, ed., *The Idea of Race in Latin America, 1870–1940* (Austin: University of Texas Press, 1990); Michael Hanchard, *Orpheus and Power: The Movimento Negro of Rio de Janeiro and Sao Paulo, 1945–1988* (Princeton: Princeton University Press, 1994); NACLA, "The Black Americas, 1492–1992," *Report on the Americas* 25 (February 1992); Peter Wade, *Race and Ethnicity in Latin America* (London: Pluto Press, 1997); and Anthony W. Marx, *Making Race and Nation* (Cambridge: Cambridge University Press, 1998).

predictably to institutional change. As this article highlights, however, one cannot isolate these governmental institutions from their moorings in state and society. Indeed, we need to *analyze* democratic politics in the context of *state-society relations* by evaluating the reach of state institutions and assessing the broader social forces that surround, support, and oppose the terms of democracies' new institutions. Failing to do so, we necessarily underestimate the extent to which the third wave has initiated a new period of democratic politics in which the social, political, and institutional terms of political exchange remain qualitatively more open and more uncertain than depicted by the contemporary institutional theorizing about democratic consolidation.

Citizenship Regimes and the Uneven Reach of the State

The contemporary round of democratization restored electoral politics to Latin America, reviving and reforming democratic rule throughout the region. These political regime changes did more than usher in new electoral institutions, however. Indeed, they oversaw a radical shift in the content of citizenship that set state-society relations along a new course. Marshall has underscored that citizenship is a differentiated bundle of rights and responsibilities that can include civil rights (freedom of organization and expression), political rights (suffrage), and social rights (the right to a minimum standard of living).[5] The contemporary democracies in Latin America have reshuffled this trilogy of potential citizenship rights—in some places expanding who may vote and in all places restricting the social rights and responsibilities that citizenship confers on citizens and states alike. In Latin America's new democracies the changing content of citizenship has also been accompanied by a change in the primary modes of interest intermediation between state and society, such that the new regimes have significantly weakened corporatist institutions and tried to replace them with more pluralist forms.

The patterned combination of citizenship rights and the accompanying modes of interest intermediation is referred to here as "citizen-

[5] T. H. Marshall, "Citizenship and Social Class," *Class, Citizenship, and Social Development* (Garden City, N.Y.: Doubleday, 1963). This article does not assume, as does Marshall, that there is a natural progression of citizenship rights and/or that they are irreversible. For alternative discussions of citizenship, see Ronald Beiner, ed., *Theorizing Citizenship* (Albany: State University of New York Press, 1995); Will Kymlicka, *Multicultural Citizenship* (Oxford: Oxford University Press, 1995); Yasemin Nuhoğlu Soysal, *Limits of Citizenship: Migrants and Postnational Membership in Europe* (Chicago: University of Chicago Press, 1994); and Gershon Shafir, ed., *The Citizenship Debates: A Reader* (Minneapolis: University of Minnesota Press, 1998).

ship regimes."[6] Empirically speaking, modes of interest intermediation and citizenship rights have tended to go together, even though theoretically speaking they do not have to. These citizenship regimes have taken different forms over time. Thus, from the mid-twentieth century on, Latin American states, whether democratic or authoritarian, tended to promote corporatist citizenship regimes. They extended social rights (including subsidies, credit, health care, education, and the like) and institutionalized corporatist modes of interest intermediation for workers and peasants, in particular. In the new democracies, by contrast, states have tended to promote neoliberal citizenship regimes.[7] The expansion of political and civil rights has tended to coincide with the decline in social rights and the promotion of liberal or pluralist modes of interest intermediation. Organized social sectors (such as workers and peasants) have lost their state assurance of a basic standard of living and similarly have lost their main institutional means of accessing and occasionally influencing the state. Although citizenship regimes have such significant consequences for state-society relations, however, they are neither equal to nor derivative of political regimes. As we will see, corporatist and neoliberal citizenship regimes developed in democratic and authoritarian regimes in Latin America (and Western Europe, for that matter).

Both corporatist and neoliberal citizenship regimes profoundly and intentionally reshaped state institutions and resources, as well as the terms of public access to them. Because of the uneven reach of the state, however, they had unintended consequences. Thus, in attempting to restructure society into class-based federations that could be controlled from above, corporatist citizenship regimes unwittingly provided autonomous spaces that could shelter rural indigenous communities from state control. And for their part, neoliberal citizen-

[6] I borrow the phrase "citizenship regime" from Jenson and Phillips. They use the term to refer to the varying bundles of rights and responsibilities that citizenship can confer. This article expands the term to refer not only to the content of citizenship but also to its accompanying modes of interest intermediation. See Jane Jenson and Susan D. Phillips, "Regime Shift: New Citizenship Practices in Canada," *International Journal of Canadian Studies* 14 (Fall 1996).

[7] I label the latter citizenship regime as *neo*liberal for three reasons. First, I want to distinguish it from T. H. Marshall's description of earlier British liberal citizenship regimes, where civil and political rights were extended but social rights were not yet on the political agenda. The sequencing of citizenship rights that Marshall identified, while perhaps applicable to the late-nineteenth-century liberal periods in Latin America, does not apply to the contemporary Latin American context, where social rights were dismantled and civil and political rights extended. Second, I want to distinguish it from the liberal periods that marked the second half of nineteenth-century Latin American politics. Finally, I want to link the contemporary neoliberal citizenship regimes to the contemporary neoliberal reforms that have redefined Latin America's political economies and dismantled many of the social programs that were once tied to social rights.

ship regimes setting out to shatter corporatism's class-based integration and replace it with a more atomized or individuated set of state-society relations in fact challenged the indigenous local autonomy that corporatism had unknowingly fostered, failed to secure the individual rights that neoliberalism had promised, and consequently politicized ethnic cleavages throughout the region. In short corporatist and neoliberal citizenship regimes had foundational projects for state and society that were consequential but unevenly institutionalized. From the top looking down, these projects restructured society in radical ways. From the bottom looking up, however, these new projects of state formation and interest intermediation have been contested at many steps along the way. This section juxtaposes the formal goals and the unintended consequences of these two citizenship regimes.

Corporatist Citizenship Regimes

At midcentury most Latin American states were experimenting with a corporatist form of citizenship regime.[8] Latin American corporatism did not necessarily advance political rights (suffrage), for it was constitutive as much of authoritarian regimes as of democratic ones. However, it generally advanced the idea that citizens have some civil rights (the right to organize under certain circumstances) and some social rights (the right to a basic standard of living). It created and/or promoted labor and peasant associations that (1) structured and often monopolized official representation, (2) received state subsidies, and (3) were controlled by the state. As Collier notes, however, the degree to which corporatism actually structured, subsidized, and controlled these federations varied significantly from case to case and over time.[9]

As part of this mid-twentieth-century corporatist project Latin American states incorporated Indians. They sought to cast aside ethnic categories (which supported ongoing attempts at nation building) and to reconstitute Indians as national peasants.[10] The states did so largely

[8] For classic perspectives on Latin American corporatism, see James M. Malloy, ed., *Authoritarianism and Corporatism in Latin America* (Pittsburgh, Pa.: University of Pittsburgh Press, 1977). For a seminal comparative analysis of Latin American corporatism, see Ruth Berins Collier and David Collier, *Shaping the Political Arena: Critical Junctures, the Labor Movement, and Regime Dynamics* (Princeton: Princeton University Press, 1991).

[9] David Collier, "Trajectory of a Concept: 'Corporatism' in the Study of Latin American Politics," in Peter H. Smith, ed., *Latin America in Comparative Perspective: New Approaches to Methods and Analysis* (Boulder, Colo.: Westview Press, 1995). State commitments to class-based federations, for example, weakened under the authoritarian regimes of the 1970s.

[10] Ethnicity and class are not the only axes for organizing grassroots mobilization in rural Latin America. Political parties, religious organizations, and cooperatives, for example, also competed for membership. However, from a national and comparative perspective they were generally not the most important players in redefining the rural landscape.

through land reforms that "emancipated" Indians from repressive and/or exploitative forms of labor control (thereby holding out to them the prospect of autonomous citizenship), occasionally distributed land and credit (thereby extending social rights), and incorporated them through peasant associations (thereby organizing them along corporatist lines). Land reforms in Mexico (1934), Bolivia (1953), Guatemala (the short-lived reform of 1952), Ecuador (1964 and 1973), and Peru (1968), for example, weakened landed elites' control of the countryside, redistributed significant tracts of land, and provided incentives for Indians to register as peasant communities. This registration reorganized the countryside along state-regulated corporatist lines, with many peasant communities joining peasant federations in hopes of gaining access to land and the state. These corporatist reforms brought with them the creation and expansion of social services in the areas of agricultural support, infrastructure, education, and health. Access to land and these services was often gained through corporatist associations. In short, the state and union organizations imposed a class identity on Indians as the ticket for political incorporation and access to resources.

The registration of peasant communities and the growth of peasant federations, in particular, fostered the fiction that the state had turned Indians into peasants and stripped indigenous ethnicity of its salience. Until recently, studies of corporatism highlighted the strong reach of these corporatist institutions and their capacity to control and remake these social sectors. Latin American corporatist states presumably centralized state-society relations. Yet this enterprise was compromised by the absence of a rationalized bureaucracy, the failure to establish authority, and a lack of monopoly on the legitimate use of force. Hence, despite official statements and institutions of corporatist control, large areas of the country operated beyond the reach of the state. Authoritarian enclaves were dominated by patronage and clientelist networks. Caudillos and landlords at times deployed their own paramilitary forces, created their own political rules, displayed greater allegiance to subnational politics than to national politics, and/or deployed state institutions for their benefit.[11] Moreover, studies of the Amazon have long noted the failure of states to govern the Amazon—leaving large

[11] Jonathan Fox, "The Difficult Transition from Clientelism to Citizenship: Lessons from Mexico," *World Politics* 46 (January 1994); idem, "Latin America's Emerging Local Politics," *Journal of Democracy* 5 (April 1994); Frances Hagopian, *Traditional Politics and Regime Change in Brazil* (New York: Cambridge University Press, 1996); Gilbert M. Joseph and Daniel Nugent, eds., *Everyday Forms of State Formation: Revolution and the Negotiation of Rule in Modern Mexico* (Durham, N.C.: Duke University Press, 1994); and R. Andrew Nickson, *Local Government in Latin America* (Boulder, Colo.: Lynne Rienner, 1995).

swaths of territory and significant numbers of Indians beyond the political and military control of the state.[12] In short, the uneven reach of the state undermined the centralizing program.

In this context, state efforts to build and register peasant communities had unintended consequences. Via land reform and credit programs, Indians secured the spaces in which they could institutionalize indigenous community practices at the local level.[13] In more ways than one, the distribution of inviolable communal lands to registered peasant communities provided Indians with the physical space not only for farming but also for securing governance by traditional indigenous authorities. In this way the legal registration of communities and granting of community-based property created a legally defined, state-sanctioned geographic area that allowed for the growth and/or maintenance of politically autonomous local enclaves, indigenous culture, and political practices. Otherwise stated, land reforms masked the maintenance of indigenous autonomy and often engendered the (re)emergence of indigenous leaders, the (re)constitution of communities, and the expression of (evolving) indigenous identities at the community levels.

In Mexico, for example, the land reform accompanied the creation of a national peasant federation, the CNC, and distributed property in many forms. Of these, the distribution of *ejidos* (communally owned land) unwittingly provided the greatest latitude for local indigenous autonomy—they were community based, inalienable, and, while regulated, often beyond state control.[14] In Bolivia the national revolutionary governments of the 1950s and the subsequent military governments between 1964 and 1974 also incorporated Indians into the state as peasants. As in Mexico, they depended on alliances and pacts with

[12] States did not actively seek to harness the Amazon region until the latter part of the twentieth century. Prior to that they had mapped out boundaries that de facto included Indians as members, though not necessarily citizens, of the given state. See Lucy Ruiz, ed., *Amazonía: Escenarios y conflictos* (Quito: CEDIME and Ediciones Abya Yala, 1993); Fernando Santos Granero, ed., *Globalización y cambio en la amazonía indígena* (Quito: FLACSO and Ediciones Abya Yala, 1996); Richard Chase Smith, "La política de la diversidad. COICA y las federaciones étnicas de la Amazonía," in Stefano Varese, ed., *Pueblos indios, soberanía y globalismo* (Quito: Ediciones Abya Yala, 1996).

[13] In Eugen Weber's classic study of nation building, he illuminates how the French state turned peasants into Frenchman. See Weber, *Peasants into Frenchmen: The Modernization of Rural France, 1870–1914* (Stanford, Calif.: Stanford University Press, 1976). I suggest here that Latin American efforts to turn Indians into peasants in fact created the space in which they could defend and develop a local indigenous identity.

[14] Jeffrey W. Rubin, *Decentering the Regime: Ethnicity, Radicalism, and Democracy in Juchitán Mexico* (Durham, N.C.: Duke University Press, 1997); Neil Harvey, *The Chiapas Rebellion* (Durham, N.C.: Duke University Press, 1998); Valentina Napolitano and Xochitl Leyva Solano, eds., *Encuentros Antropológicos: Politics, Identity, and Mobility in Mexican Society* (London: Institute of Latin American Studies, 1998); Fox (fn. 11); and Shannan Mattiace, "!Zapata Vive! The Ezln, Indian Politics and the Autonomy Movement in Mexico," in George A. Collier and Lynn Stephen, special eds., *Journal of Latin American Anthropology* 3, no. 1 (1997).

peasant federations, which were expected both to deliver votes to tl government and to control the local communities. Contrary to tl hopes of politicians and military officers, Bolivia simultaneously wi nessed the defense and/or reconstitution of *ayllus* (kinship groups go erned by a set of local-level indigenous authorities).[15] In Ecuador tl 1937 community law and later the 1964 and 1973 land reforms defin indigenous men and women as peasants and gave them access to tl state insofar as they represented themselves as peasant communiti and/or unions. Indeed, the number of registered peasant communiti skyrocketed in the 1960s and 1970s.[16] However, at the local level, mai indigenous communities continued to maintain some form of indig nous practices and institutions.[17]

Corporatism, therefore, created a dynamic dualism, with identiti shifting according to the locale: for the state, Indians assumed identiti as peasants; within the community, peasants assumed their identities Indians.[18] Location therefore mattered for the expression of identit Where the state incompletely penetrated local communities (nowhe more evident than in the Amazon), Indians sustained a certain degr of political autonomy by retaining and/or creating authority systen and customs.[19] Over time the boundaries did not remain so clear. Ii digenous authorities and rules shaped union politics just as union ai thorities and rules began to shape community dynamics.

NEOLIBERAL CITIZENSHIP REGIMES

With the transition from authoritarian rule in the late 1970s and tl early 1980s, Latin American reformers reestablished democratic inst

[15] Silvia Rivera Cusicanqui y equipo THOA, *Ayllus y proyectos de desarrollo en el norte de Potosí* (La P Ediciones Aruwiyiri, 1992); Esteban Ticona A., Gonzalo Rojas O., and Xavier Albó C., *Votos y wipl las: Campesinos y pueblos orginarios en democracia* (La Paz: Fundación Milenio and CIPCA, 1995); and J liana Ströbele-Gregor, "Culture and Political Practice of the Aymara and Quechua in Boliv Autonomous Forms of Modernity in the Andes," *Latin American Perspectives* 23 (Spring 1996).

[16] Leon Zamosc, *Estadística de las áreas de predominio étnico de la sierra ecuatoriana: Población rural, i dicadores cantonales y organizaciones de base* (Quito: Ediciones Abya Yala, 1995).

[17] A similar pattern emerged following land-reform programs in Guatemala (1944–54) and Pe (1968). Given high levels of repression, however, corporatist policies and institutions were undermin and dismantled shortly after they were created. Nonetheless, the general outline of this argument i mains. While states promoted national ideals, indigenous communities found ways to shelter th right to sustain and develop ethnic identities and ties.

[18] This duality is captured by disciplinary differences in the social sciences. Political scientists wor ing on this period have highlighted the centrality of class, the peasantry, and corporatist organizatio as if they displaced community autonomy and ethnic identities. Anthropologists have historically f cused on the local level and, in turn, have highlighted community autonomy and ethnicity, often at t expense of broader patterns of state-society relations.

[19] Corporatist citizenship regimes barely penetrated the Amazon. Amazonian Indians rarely form part of peasant federations and states did not have the resources to control them. Consequently, Am zonian Indians had even more autonomy than Andean and Mesoamerican Indians.

tutions. The original democratic reforms left intact many of the vertical corporatist institutions and corporatist social policies just discussed. However, with the economic crisis of the 1980s and 1990s, many states began to reassess their capacity for maintaining a corporatist citizenship regime and the desirability of doing so, with all that that entailed—social rights plus structured and hierarchical class-based federations that were to mediate between state and society. Adopting a neoliberal discourse, they advocated instead individual autonomy and responsibility, a program based on granting individual political and civil rights (but not necessarily social rights), emasculating corporatist organizations, and advocating the retreat of the state.[20]

This shift in citizenship regimes has had significant consequences for indigenous peoples. Individual rights have been promoted at the expense of corporate organizations—peasant and labor federations—that had served as the primary (at times the only) mode of interest intermediation between state and society. These federations have lost political and social leverage throughout the region, and with this Indians have lost their formal ties to the state.[21] The shift has also informed the adoption of stabilization and structural adjustment policies that have drastically cut back social services and goods that ostensibly were designed to secure a basic social standard of living for citizens. Most dramatically for Indians, states have privatized land markets, liberalized agricultural prices, eliminated agricultural subsidies, and diminished credit programs.[22] These reforms (particularly efforts to privatize land markets and to privilege the individual over the corporate unit) echo late-nineteenth-century liberal reforms that were incontrovertibly detrimental to indigenous peoples. In both cases, the reforms threatened a communal land base that the state had once made inviolable.[23]

[20] There is a healthy and unresolved theoretical debate about whether the core of liberal thought is based on toleration or autonomy. But it would be foolish to argue in the Latin American context that there is one coherent core, as the region's history of liberalism is undeniably syncretic.

[21] Several states did have national indigenous institutes. However, these rarely if ever served as interlocutors between Indians and the state.

[22] Catherine M. Conaghan and James M. Malloy, *Unsettling Statecraft: Democracy and Neoliberalism in the Central Andes* (Pittsburgh, Pa.: University of Pittsburgh Press, 1994); Miguel Urioste Fernández de Córdova, *Fortalecer las comunidades: Una utopía subversiva, democrática . . . y posible* (La Paz: AIPE/PROCOM/TIERRA, 1992); Nora Lustig, *Coping with Austerity: Poverty and Inequality in Latin America* (Washington, D.C.: Brookings Institution, 1995); Alain de Janvry et al., *The Political Feasibility of Adjustment in Ecuador and Venezuela* (Paris: OECD, 1994); Samuel A. Morley, *Poverty and Inequality in Latin America: The Impact of Adjustment and Recovery in the 1980s* (Baltimore and London: Johns Hopkins University Press, 1995).

[23] Indians resisted both sets of reforms. However, the isolated and ephemeral terms of indigenous historical resistance differ from the more organized and sustained contemporary indigenous movements. See fn. 2.

The reconfiguration of local political power, control, and administration has also occurred through decentralization programs to check the central state and to break up the concentration of administrative, fiscal, and political power. Although changes in Bolivia, Brazil, Colombia, Guatemala, and Venezuela have been particularly noteworthy, the new trends are evident throughout the region.[24] Decentralization programs, as a new type of state formation, have devolved varying degrees of power to more local units (municipalities, provinces, regions, and so on) and, in turn, have created a more localized space for agglomerating individual preferences, calculating decisions, and implementing programs. While creating a new layer of political bureaucracy, they have also created additional entry points for individuals to try to shape local and national politics. Moreover decentralization has created more possibilities for holding political leaders accountable for their (in)action. It has attempted to institutionalize a new form of democratic politics: both more liberal and more local. The geography of democracy, in other words, has shifted away from rationally sanctioned corporatist institutions and toward more pluralist and local forms of interest intermediation.

With the promise of equal rights and greater political participation, neoliberal citizenship regimes would seem to hold out great hope for democratizing politics. Yet, as the next section underscores, Latin American states have not secured these neoliberal citizenship rights—in particular, for Indians—even while they start to take away the social rights, access to the state, and indigenous local autonomy once unwittingly associated with corporatist citizenship regimes. This inability to secure neoliberal citizenship regimes is in large part a function of the weak reach and retreat of the state.

Weber argued in his classic statement: "The state is a human community that (successfully) claims the monopoly of the legitimate use of physical force within a given territory."[25] In Latin America, however, as in most of Africa, Asia, and the Middle East, this standard is still largely unmet. Many of Latin America's central political institutions remain weak, commitment to those institutions remains questionable, and the territorial scope of those institutions remains ambiguous.[26]

[24] Eliza Willis, Christopher da C.B. Garman, and Stephan Haggard, "The Politics of Decentralization in Latin America," *Latin American Research Review* 34, no. 1 (1999).

[25] Max Weber, "Politics as a Vocation," *From Max Weber: Essays in Sociology*, ed. and trans. H. H. Gerth and C. Wright Mills (New York: Oxford University Press, 1946), 78.

[26] State formation is a process of political mapping. As Scott has argued, it requires a situation of mutual intelligibility. The state must be able to read, identify, and defend the territory it governs. Those governed should be able to identify (with) and depend on the state for basic functions. See James C. Scott, *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed* (New Haven: Yale University Press, 1998). See also fn 1.

This is nowhere more apparent than from the vantage of the countryside. From that perspective, it is difficult to argue that there is a single human community (as opposed to many), that the state claims a monopoly on the legitimate use of force, or that the territory is clearly defined. National identities, borders, and legitimacy are all in question and often in flux. Indeed, Latin America remains very much in the throes of state formation, where the identities, borders, and legitimacy of the state are highly politicized and contested processes, particularly in the countryside. It is not surprising therefore that consolidating democracy and securing neoliberal citizenship regimes under these conditions remains a tall order.

In this context ethnic cleavages have been politicized and indigenous movements—particularly those that mobilize in the countryside—have protested the ongoing violation of their political rights under democratic regimes, with significant political consequences.[27] As stated most boldly by CONAIE, Ecuador's largest and most prominent indigenous movement: "In Ecuador the fundamental principles of democracy—equality, liberty, fraternity, and social peace—have not been achieved."[28] Throughout the region indigenous leaders have made similar points. They recount the ways in which individuals' rights have been dismissed—in voting booths, courts, and schools—and argue that the state should do more to uphold and protect their individual rights.[29] Hence, despite a discourse of individual civil and political rights, states remain incapable of protecting them.[30] The state's inability to secure individual rights makes many indigenous communities even more wary of the restrictions that neoliberal citizenship regimes would place on the inalienable *community rights* and de facto *local autonomy* that they had secured during the prior corporatist citizenship regime.

[27] This article seeks to explain not why indigenous movements *emerge* but why indigenous identities became *politicized* and how that politicization is reflected in the postliberal challenge, a regionwide agenda for democratic reform. For an explanation of why, when, and where indigenous organizations emerge in Latin America, see Yashar (fn. 2, 1998); indigenous movements emerge only where state reforms that challenge local autonomy combine with political liberalization and preexisting networks. In some cases these movements were supported and shaped by ties to urban and international actors.

[28] CONAIE, *Proyecto político de la CONAIE* (Quito: CONAIE, n.d.), 6.

[29] Based on anonymous small-group discussions conducted by the author in Ecuador, Bolivia, and Peru during the course of 1997, and repeated in most 1995–97 interviews with indigenous leaders in Ecuador, Bolivia, Peru, Guatemala, and Mexico.

[30] Rodolfo Stavenhagen and Diego Iturralde, eds., *Entre la ley y la costumbre: El derecho consuetudinario indígena en América Latina* (Mexico City: Instituto Indigenista Interamericano; San José, Costa Rica: Instituto Interamericano de Derechos Humanos, 1990); Jorge E. Dandler, "Indigenous Peoples and the Rule of Law in Latin America: Do They Have a Chance?" (Paper prepared for the academic workshop on the Rule of Law and the Underprivileged in Latin America, Helen Kellogg Institute for International Studies, University of Notre Dame, November 9–11, 1996); and Ramón Torres Galarza, ed., *Derechos de los pueblos indígenas: Situación jurídica y políticas de estado* (Quito: Abya Yala, CEPLAES, CONAIE, n.d.).

THE POSTLIBERAL CHALLENGE

In this context ethnic cleavages have been politicized and indigenous movements have challenged the neoliberal shift in democratic citizenship.[31] They have exposed the incomplete implementation of neoliberal citizenship regimes and have proposed a more complex political geography for defining units and locales of democratic representation and governance. Yes, they want more liberal politics insofar as the state would defend individual rights that in practice have been denied. Yes, they want more local power insofar as it would increase access to the state and defend local autonomy. However, they do not uniformly subscribe to the "more liberal more local" formula. Indeed, they question the state's capacity to uphold neoliberal citizenship rights just as they challenge the universalizing assumptions of neoliberal citizenship regimes about national identity, unit representation, and state structure.[32]

In this context, indigenous movements now pose a postliberal challenge, by demanding a different kind of political mapping—one that would secure individual rights but also accommodate more diverse identities, units of representation, and state structures.[33] What follows is a stylized presentation of this postliberal challenge. This section addresses the content of the postliberal demands that Indian movements have placed on the political agenda. This is only an overview, however, and not a full description of the complex and diverse indigenous demands throughout the region. It is meant to be illustrative of the highly contested terms of contemporary politics in the region.

CHALLENGING NATIONAL HOMOGENEITY

The Latin American indigenous movements that have emerged with this third wave of democratization have started to challenge the project of nation building and assimilation that was associated with nineteenth-century liberal parties and that has been inscribed in Latin

[31] Ethnic cleavages were also politicized in Nicaragua in the 1980s, as the Miskitu demanded, struggled for, and achieved autonomy. This politicization occurred, however, during the revolutionary decade headed by the Sandinistas—a very different historical and political context from the shifting citizenship regimes in the rest of the region. See Charles R. Hale, *Resistance and Contradiction: Miskitu Indians and the Nicaraguan State, 1894–1987* (Stanford, Calif.: Stanford University Press, 1994).

[32] See Iris Marion Young, "Polity and Group Difference: A Critique of the Ideal of Universal Citizenship," in Beiner (fn. 5). Young argues that liberal democracies profess to represent all individuals equally but, in fact, privilege certain dominant voices over others. She calls for a differentiated form of citizenship, one in which social groups are granted spaces for representation, participation, and voice. While indigenous peoples have not necessarily read Young, their claims in fact parallel hers when they indicate that Indians should gain additional and different rights alongside individual ones.

[33] Parallel theoretical discussions about politics in advanced industrial democracies have tended to refer to this challenge as the "multicultural" challenge.

merican constitutions since then. Nineteenth-century liberals en-
gged in nation-building projects that sought to create national unity—
policy that legitimated the assimilation of indigenous peoples and
tacks on indigenous communal lands. Building on nineteenth-cen-
ry liberal ideas, Latin American constitutions during both corporatist
nd neoliberal citizenship regimes have codified national identities as
e basis for political membership. Confronted with diverse ethnic
pulations, states have promoted assimilation—even while they main-
ined labor markets that were segmented along ethnic lines—to
hieve this national ideal. Where nationals did not exist, they would
rge them. In this regard, national politicians and constitutions have
ther assumed ethnic homogeneity or disregarded the political salience
ethnic diversity.

Contemporary indigenous movements challenge both claims. Their
ry presence undermines claims of ethnic universality. Indeed, data on
digenous populations highlight how unsuccessful assimilation and
ation-building projects have been in the long term. Certainly indige-
ous populations have suffered a decline in absolute and relative terms.
dians, however, still constitute a substantial portion of the national
pulations in Bolivia (71.2 percent), Ecuador (37.5 percent),
uatemala (60.3 percent), Mexico (12.4 percent), and Peru (38.6 per-
nt), with smaller percentages in the rest of the region.[34] Indeed, these
rcentages are testament to the failure of nation-building projects
nce the nineteenth century and to the capacity of indigenous commu-
ties to maintain indigenous community structures and identities at
e local level. As discussed, these identities were nursed and developed
the local level, even during the corporatist period that attempted to
nvert Indians into peasant nationals.

With democratization and the turn to neoliberal citizenship
gimes, indigenous movements in Mexico, Guatemala, Colombia, Bo-
ia, Ecuador, and Brazil have demanded constitutional reforms recog-
zing the multiethnic and plurinational composition of their countries.
hese demands highlight the endurance of many ethnic communities
ven while the content of those identities has surely changed) despite
tion-building projects. As part of this effort, indigenous movements
ve appealed to norms, laws, and organizations operating in the inter-
tional arena.[35] In particular, they have lobbied Latin American states

[34] Yashar (fn. 2, 1996). These figures should be read with caution given the problems of data collec-
n and measurement.

[35] Alison Brysk, "Acting Globally: Indian Rights and International Politics in Latin America," in
n Cott (fn. 2); idem, "Turning Weakness into Strength: The Internationalization of Indian Rights,"

to ratify the International Labour Organization (ILO) Convention on Indigenous and Tribal Peoples in Independent Countries. Conv tion 169 outlines the rights of indigenous peoples and the responsi ities of multiethnic states toward them. At a minimum, it calls on st to recognize ethnic heterogeneity where states had advanced natio ist aspirations of mestizo homogeneity. The following Latin Ameri states have ratified ILO Convention169: Mexico (1990), Bolivia (19 Colombia (1991), Costa Rica (1993), Peru (1994), Paraguay (19 Honduras (1995), Guatemala (1996), and Ecuador (1998).[36] Ratifi tion provides a mechanism for advocating constitutional reforms to commodate ethnically diverse populations; it should not necessaril seen as a prelude to secession.[37]

While these Latin American states have yet to live up to the term the convention, they are beginning to discuss constitutional ame ments that recognize the multiethnic and pluricultural makeup of e country, as in Mexico (1992), Bolivia (1994), and Ecuador (1998 These reforms are an important symbolic victory for indigenous p ples who have worked to change myths of national unity. Indeed, constitutional recognition of ethnic heterogeneity in some Latin Am ican states has opened up possibilities to discuss and debate other ki of democratic institutions that can accommodate a diverse ethnic p ulation, including discussions about consociationalism, identity-ba electoral institutions, ethnic political parties, federalism, and the lik

CHALLENGING UNIT HOMOGENEITY[40]

This demand for multicultural recognition is the first step toward m ing claims that Indian cultures cannot be reduced to individual ide ties and rights, as neoliberal citizenship regimes would have it, bu

Latin American Perspectives 23, no. 2 (1996); and Franke Wilmer, *The Indigenous Voice in World P* (Thousand Oaks, Calif.: Sage Press, 1993). As these authors note, the international arena has pro a new discourse, funds, and forums that have often shaped debates about indigenous ri

[36] http://ilolex.ilo.ch:1567/scripts/ratifce.pl?C169. Thirteen countries have ratified Convention including Denmark, Fiji, the Netherlands, and Norway.

[37] The call for constitutional recognition of multiethnic and plurinational populations has eli vitriolic reactions from some politicians. Specifically, they fear that this recognition of different " ples" will provide Indians with the leverage to appeal to UN laws that sanction the right of all pe to self-determination and, by implication, to their own state.

[38] Dandler (fn. 30). In a May 1999 Guatemalan referendum the voting population (18 perce the eligible electorate) rejected these proposed reforms.

[39] See Arend Lijphart, *Democracy in Plural Societies: A Comparative Exploration* (New Haven: University Press, 1977); Hans Daalder, "The Consociational Democracy Theme," *World Politi* (July 1974); Ian S. Lustick, "Lijphart, Lakatos, and Consociationalism," *World Politics* 50 (Oct 1997); and Horowitz (fn. 3).

[40] "Unit homogeneity" refers here to the unit of political representation and intermediation. It i meant to evoke the standard meaning of the term used by methodologists.

fact rest on primary and collective sets of identities, organizations, and ·ights. Accordingly, indigenous movements have started to challenge :he effort to homogenize and individualize the appropriate unit of poitical representation and intermediation.

Latin American constitutions have tended to assume unit homogeneity. As noted, whereas Latin American regimes once privileged :orporatist forms of interest intermediation, they now privilege individuals as the primary unit. With the current neoliberal citizenship ·egimes, Latin American politicians have contended that the central political unit is and should be the individual. The individual chooses to vote, to join political parties, to participate in organizations, and to hold government accountable. In short, the individual is the foundational ınit of rights and responsibilities in a polity presumed to be moving toward a more liberal democracy. Policymakers have voiced concern ıbout equalizing treatment before a state that engaged in indiscrimiıate repression and torture and, to that end, advocate paying closer atention to the rule of law. In a context where dissidents were killed or ailed, indigenous people were excluded and/or repressed, and regions vere controlled by local bosses, the call for a universalizing set of norms ınd institutions to protect individuals is an important normative step oward deepening democracy.

Given the democratizing intentions of neoliberal citizenship 'egimes, it is striking that indigenous men and women are cautious ıbout the drive to promote the individual as the primary political unit of democracy. Where the dominant political discourse suggests the ad'ance of individual rights, indigenous communities often foresee an infringement on indigenous autonomy and resources. Because indigenous :ommunities have rarely experienced the full complement of civil and political rights associated with liberal democracy, they have little reason o believe that neoliberal citizenship regimes will necessarily fulfill their promises now. To the contrary, they see the promotion of neoliberal versus corporatist) citizenship regimes as an infringement on social ights that once enabled them to act as autonomous communities.

Confronted with this shift in citizenship regimes, contemporary inligenous movements have brought into sharp relief the tense interplay ·etween the contemporary celebration of the individual and indigenous ommunity practices.[41] Indigenous movements generally argue that the

[41] See Enrique Mayer, "Reflexiones sobre los derechos individuales y colectivos: Los derechos étni›s," and Rodolfo Stavenhagen, "Los derechos indígenas: Algunos problemas conceptuales," both in ric Hershberg and Elizabeth Jelín, eds., *Construir la democracia: Derechos humanos, ciudadanía y soedad en América Latina* (Caracas: Nueva Sociedad, 1996).

individual should not be the only unit of representation, nor should it be privileged. They demand that the state uphold equal rights and responsibilities for Indians as individuals and in this sense are calling for the fulfillment of liberal ideals. But they argue as well that the state should recognize indigenous communities as a historically prior and autonomous sphere of political rights, jurisdiction, and autonomy. These demands range from the call for community or supracommunity autonomy to the call for designated representation in legislatures and ministries. The new Colombian constitution, for example, has allocated two seats in the national legislature for indigenous representatives.

If we look at a series of indigenous movements in Mexico (EZLN), Guatemala (COMG), Ecuador (CONAIE), Bolivia (CIDOB and CSUTCB), and Peru (AIDESEP), we find that their strategies have differed. In some cases they have taken up arms (Mexico); in others they have organized marches for recognition (Ecuador and Bolivia); in others they have negotiated directly with the government for new laws that recognize communities (Mexico, Guatemala, Bolivia); and in others they have used existing administrative laws that map out local political units to secure a de facto space in which indigenous communities can indirectly act as a political unit (Peru). But despite these differences, we find each movement demanding that the state simultaneously protect members' individual civil and political rights *and* recognize indigenous communities as a political unit. This position is forcefully articulated in several movement documents,[42] as well as in interviews with leaders from each of the movements.[43] At present this issue of indigenous peoples and their political representation/participation as individuals versus communities is being hotly debated and negotiated throughout the region.[44]

[42] CONAIE (fn. 28), 11–12; COMG, *Construyendo un futuro para nuestro pasado: Derechos del pueblo maya y el proceso de paz* (Guatemala City: Editorial Cholsamaj, 1995); Servicios del Pueblo Mixe, A.C., "Autonomía, una forma concreta de ejercicio del derecho a la libre determinación y sus alcances," *Chiapas 2* (Mexico City: Ediciones ERA, S.A. de C.V., 1996).

[43] These statements were made in 1997 author interviews with indigenous leaders: from Peru, with Evaristo Nukguaj and Bonifacio Cruz Alanguía; from Bolivia, with Marcial Fabricano and Román Loayza; and from Ecuador, with Luis Macas, Leonardo Viteri, César Cerdas, and Valerio Grefa. Similar statements were made in an author interview in the United States with Guatemalans Manuela Alvarado, Alberto Mazariegos, and Juanita Bazibal Tujal, May 3, 1998.

[44] For normative debates in political science on the topic of individual and group rights, see Will Kymlicka, ed., *The Rights of Minority Cultures* (Oxford: Oxford University Press, 1995); and Ian Shapiro and Will Kymlicka, eds., *Ethnicity and Group Rights* 39, Yearbook of the American Society for Political and Legal Philosophy (New York: New York University Press, 1997). Democratization studies of Latin America have generally ignored the liberal-communitarian debate that speaks to the philosophical foundations of ethnic diversity and democratic representation. These debates, however, raise the question of the central unit of political life. They have also discussed whether in fact communities should be granted special (i.e., different) rights by virtue of being a community. There have been efforts to conjoin these seemingly opposite positions, for example, by Kymlicka, "Introduction," and Kukathas, "Are There Any Cultural Rights?" in Kymlicka, *The Rights of Minority Cultures*. The author

CHALLENGING ADMINISTRATIVE HOMOGENEITY

Many indigenous movements have also started to challenge basic normative assumptions about the desirability of universal administrative boundaries. Rejecting state-formation projects that have sought to centralize or decentralize political institutions according to a single blueprint, indigenous movements throughout the region have demanded that the state recognize administrative boundaries that are unique to indigenous peoples. In this regard, it is not enough to promote municipalization, decentralization, and accountability as a means of increasing the representation, accountability, and transparency of local governments. To the contrary, indigenous peoples are increasingly demanding that the state recognize territorial boundaries (even, or particularly, where they cut across municipal or provincial boundaries) in which social relations are regulated by indigenous authority systems and customary law. In other words, they are arguing that a differentiated citizenship should coincide with differentiated administrative boundaries.[45]

Demands for territorial autonomy are growing throughout the Amazon. Although the state historically had been weak there, it is currently seeking to penetrate more deeply into these areas, privatize land markets, and regulate social relations. In the 1991 Colombian Constituent Assembly, for example, Indians negotiated reforms that granted territorial autonomy.[46] In Ecuador OPIP placed territorial demands on the political map with the thirteen-day, two-thousand-person march from Puyo to Quito in 1992. The government eventually conceded 19 different territorial blocs that totaled 138 legally recognized communities and 1,115,000 hectares.[47] In Bolivia the main Amazonian indigenous

---

fail, however, to provide guidelines for how to institutionalize both individual liberal rights and communitarian rights. Because they fail to problematize sufficiently the role that the state plays in balancing these goals— particularly when communities demand a form of political autonomy that includes alternative juridical and authority systems—their normative discussion resonates only loosely with the empirical cases of democratization.

[45] Young (fn. 32) introduces the term "differentiated citizenship." References to country-specific demands for differentiated administrative boundaries follow. For a comparative overview of the current state of legal pluralism and autonomy regimes, see Dandler (fn. 30); Donna Lee Van Cott, *The Friendly Liquidation of the Past: The Politics of Diversity in Latin America* (Pittsburgh, Pa.: University of Pittsburgh Press, forthcoming); idem, "Explaining Ethnic Autonomy Regimes in Latin America" (Manuscript, 1999); and Michael Addison Smith, "Indigenous Law and the Nation States of the Latin American Region" (Manuscript, University of Texas, School of Law and the Mexican Center, April 20, 1999).

[46] See Bartolomé Clavero, *Derecho indígena y cultura constitucional en América* (Mexico: Siglo Veintiuno, 1994), 187–89; Van Cott (fn. 45, forthcoming and 1999); and idem, "A Political Analysis of Legal Pluralism in Bolivia and Colombia," *Journal of Latin American Studies* (forthcoming).

[47] Two author interviews each, in Ecuador, with Leonardo Viteri, César Cerdas, and Gonzalo Ortiz Crespo between February and May 1997.

organizations have also won territorial autonomy.[48] Demands were first articulated during the 1990 March for Territory and Dignity, organized by CPIB. The president responded by issuing presidential decrees that recognized four indigenous territories. In 1996 the government finally passed a new agrarian reform that provided indigenous communities with the legal basis for appealing for territorial recognition—including the right to vast expanses of land and the political autonomy of indigenous authorities. By August 1997 the state had recognized seven distinct territories totaling 2.6 million hectares, and it was processing thirty-four more demands totaling about 20 million hectares.[49]

Beyond the Amazon as well, there are now demands for state recognition of indigenous communities as politically autonomous units, challenging the hegemonic idea of administrative homogeneity. In Bolivia, in particular, there is a push to recognize, reconstitute, and/or register *ayllus* (communal kinship organizations) that dot the Andean countryside.[50] The 1997 agrarian reform law makes this recognition possible.[51] In Ecuador this public discussion is incipient, as indigenous movements and nongovernmental organizations have started to engage in dialogue and initiate projects to strengthen and/or reconstitute systems of elders that have receded in importance over the years.[52] Mexico and

[48] May–August 1997 author interviews with indigenous leaders Marcial Fabricano and Ernesto Noe of CIDOB, with researchers Zulehma Lehm and Wilder Molina at CIDDEBENI, and with lawyer Carlos Romero Bonifaz of CEJIS. See Kitula Libermann and Armando Godínez, eds., *Territorio y dignidad: Pueblos indígenas y medio ambiente en Bolivia* (La Paz: ILDIS and Editorial Nueva Sociedad, 1992); and Carlos Navia Ribera, *Reconocimiento, demarcación y control de territorios indígenas: situación y experiencias en Bolivia*, Working Paper no. 34 (Trindad, Bolivia: Centro de Investigación y Documentación para el Desarrollo del Beni, July 1996); Wilder M. Molina, "El movimiento social indígena del Beni en el contexto del proceso de consolidación de la movilización intercomunal hasta la Marcha por el Territorio y la Dignidad (1987–1990)" (Manuscript, Centro de Investigación y Documentación para el Desarrollo del Beni, Trinidad, Bolivia, 1997); and Van Cott (fn. 45, forthcoming and 1999, and fn. 46).

[49] Author interviews conducted in Bolivia with Isabel Lavadenz, former national director of the National Institute of Agrarian Reform, and Jorge Múñoz, researcher at UDAPSO, 1997. See also Jorge A. Múñoz and Isabel Lavadenz, "Reforming the Agrarian Reform in Bolivia" (Paper prepared for HIID/UDAPSO and presented in Cambridge, Mass., and La Paz, Bolivia, 1997).

[50] *Ayllus* often claim sovereignty over discontinuous land bases, in contrast to Western ideas of state formation that generally assume/advocate that continuous areas coincide with a single political administration.

[51] Author interviews conducted in Bolivia between May and August 1997 with former Aymaran leader Constantino Lima; Carlos Mamani, María Eugenia Choque Quispe, and Ramón Conde, researcher-activists at THOA; and Ricardo Calla, former-director of TAYPI. See Sergio Molina and Iván Arias, *De la nación clandestina a la participación popular* (La Paz: Centro de Documentación e Información CEDOIN, 1996); Xavier Albó and Ayllu Sartañani, "Participación popular en tierra de ayllus," in David Booth, ed., "Popular Participation: Democratising the State in Rural Bolivia" (Manuscript).

[52] Author interviews conducted in Ecuador between February and May 1997 with indigenous leaders José María Cabascango, Luis Maldonado, and Luis Macas.

uatemala have seen separate efforts to negotiate or proclaim autonomy r the Mayan populations residing on either side of the border.[53]

Hence, indigenous movements throughout the Americas are assert- g their right to new administrative spheres that have a certain degree 'political autonomy at the local level. This is more than just a call for ore land, although that is certainly a core component of the demand. ather, it is a demand that the state recognize indigenous political ju- sdiction over that land, including the right of indigenous legal systems id authorities to process and adjudicate claims. In this regard, diversi- ed state structures would coincide with some form of legal pluralism.

These calls might support federalism and/or decentralization but innot be reduced to either one or the other.[54] Federalism and decen- alization are designed to grant greater local sovereignty over local is- ies; federalism and decentralization assume an important degree of dministrative homogeneity. Each assumes that an entire country will e defined by federal and/or municipal administrative boundaries. Each dministrative unit (whether the state and/or the municipality) ideally overns with the same understanding of the dividing line between fed- ral/national and local jurisdiction. Many indigenous organizations upport this idea insofar as it provides additional entry points for par- icipation as both electors and elected. And indeed, with decentraliza- ion, the level of indigenous participation in elections has grown; in olivia, for example, the number of elected indigenous officials has in- reased.[55]

Demands for local autonomy, however, actually challenge the ad- ninistrative homogeneity entailed in decentralization and federalism. ndigenous organizations assert that their collective identity—which is istorically prior to the formation of each Latin American state—enti- les them to special jurisdictions that crosscut, transcend, and are dis- inct from homogenous state administrative boundaries. They want not nly more local autonomy but also more expansive jurisdiction for In-

[53] For examples of autonomy debates in Mexico and Guatemala, see *Ojarasca*, no. 45 (August–No- ember 1995); Héctor Díaz-Polanco, "La rebelión de los más pequeños: Los zapatistas y la autonomía" Manuscript); *Journal of Latin American Anthropology* 3, no. 1 (1997); the 1996 Guatemalan Peace Ac- ords; the 1995 Guatemalan *Acuerdo sobre identitad y derechos de los pueblos indígenas*; and Rachel Sieder, d., *Guatemala after the Peace Accords* (London: Institute of Latin American Studies, 1998).

[54] For the differences between territorial autonomy, federalism, and decentralization, see Ruth Lapi- oth, *Autonomy: Flexible Solutions to Ethnic Conflict* (Washington, D.C.: United States Institute of eace Press, 1996), 50–52.

[55] See Ministerio de Desarrollo Humano, Secretaría Nacional de Participación Popular, *Indígenas y el poder local* (La Paz: Talleres de Editorial Offset Boliviana, 1997); and Van Cott (fn. 45, forthcom- ig, and fn. 46).

dian communities—an arrangement that would not necessarily be accorded nonindigenous communities, whose local jurisdiction might be much more restrictive vis-à-vis the federal or national state. Hence, regardless of whether a country is defined by federal units, indigenous movements are demanding that the state recognize political and juridical spaces that are primarily occupied and administered by indigenous communities. These proposals would in fact result in a more multilayered conception of the polity, the state, and its citizens, one that would promote inclusion and autonomy simultaneously. These are not demands for secession but for institutional pluralism in multiethnic settings.

The postliberal agenda for local autonomy, however, is no panacea. Local autonomy could tend toward illiberal politics at the local level.[56] The recognition of local autonomy could provide traditional authorities with the means to carve out their own fiefdoms with few outside checks on the exercise of that power—thereby inhibiting the democratization of local life within indigenous communities. Moreover, traditional indigenous practices could disadvantage groups in society —limiting their voice, access to indigenous and nonindigenous resources, and individual autonomy—by charging that their concerns threatened the sanctity of local autonomy and tradition. Women in particular have historically been excluded from public political spheres, where the male head of household often speaks for the family unit, where women are often denied equal access to education and social services, and where battered women often have little legal recourse within the community. Consequently, the postliberal challenge could simultaneously increase local autonomy (a liberal good) and decrease local tolerance (an illiberal outcome).

In short, indigenous movements pose a postliberal challenge. They challenge the homogenizing assumptions that suggest that individuals unambiguously constitute the primary political unit and that administrative boundaries and jurisdictions should be uniformly defined throughout a country. And they call instead for more differentiated forms of citizenship and political boundaries, ones that grant individuals rights as citizens but that also grant collective rights and political autonomy at the local level. Finally, in calling for the constitutional recognition of pluriethnic and multicultural states, they challenge the idea that the state (democratic or otherwise) should correspond to a presumed homogeneous nation. In this regard, they challenge claims of

[56] *The Federalist Papers* raise parallel concerns about the power of factions in small pure democracies, particularly in the essays by James Madison; see *The Federalist Papers*, selected and ed. Roy P. Fairfield (Baltimore: Johns Hopkins University Press, 1981), no. 10.

:thnonational homogeneity and assert the political salience of ethnic liversity. By advocating a differentiated kind of citizenship, they are ›ushing to redefine democratic institutions in dramatic ways.

IMPLICATIONS FOR STUDYING DEMOCRATIC CONSOLIDATION

3y problematizing the reach of the state and the response of social 'orces to the new democratic period, this article depicts a situation in ›olitical flux. Indigenous movements and postliberal agendas pose funlamental challenges that Latin America's democracies are beginning or will need) to tackle. They are forcing Latin America's new regimes o confront the limited reach of prior rounds of state formation, to adlress the indeterminacy of the current round of democratic institution ›uilding, and to consider how new democracies might reform states .nore effectively to accommodate plural identities, political units, and administrative heterogeneity. Within this context states are debating constitutional reforms, decentralization, territorial autonomy, legal pluralism, and the like. These are critical, ongoing debates that highlight a contested and unfolding political process—one that in years to come could possibly redesign states, citizenship regimes, and, by extension, the ways in which democracy is practiced. Hence, while Latin American countries have largely made the transition to democracy, it would be difficult to argue that these regimes—and the states and societies that undergird them—are consolidated in any meaningful sense of the word.

It is striking then that a substantial wave of *theorizing* on contemporary Latin America has been so eager to explain democratic consolidation in third-wave democracies. The core of these recent studies has set out to explain failures versus successes in democratic consolidation and has focused increasingly and almost exclusively on contemporary governmental institutions.[57] "Institutions are in," as Ames has remarked.[58] These recent and increasingly dominant institutional studies assume that democratic institutions and their incentive structures can ensure regime consolidation. With government institutions as the primary

[57] The initial trend in democratic consolidation studies did not necessarily adopt this institutional focus. Notable examples include Scott Mainwaring, Guillermo O'Donnell, and J. Samuel Valenzuela, eds., *Issues in Democratic Consolidation: The New South American Democracies in Comparative Perspective* (Notre Dame, Ind.: University of Notre Dame Press, 1992); John Higley and Richard Gunther, eds., *Elites and Democratic Consolidation in Latin America and Southern Europe* (Cambridge: Cambridge University Press, 1992); and Juan J. Linz and Alfred Stepan, *Problems of Democratic Transition and Consolidation: Southern Europe, South America, and Post-Communist Europe* (Baltimore: Johns Hopkins University Press, 1996).

[58] Barry Ames, "Approaches to the Study of Institutions in Latin American Politics," *Latin American Research Review* 34, no. 1 (1999), 221.

focus, these scholars have revived important theoretical discussions and empirical comparisons of constitutional design, presidentialism versus parliamentarism, and electoral models.[59] They investigate how to tinker with government institutions in order to make them more stable and enduring. To explain democratic consolidation these comparative studies have emphasized the ways in which different governmental institutions court or accommodate political conflict. The assumption is that the same institutional model should engender either democratic consolidation or breakdown across cases and across time.

The questions posed by these authors about institutional design are politically salient; the research is increasingly systematic; and the arguments are often provocative. In this regard, there have been significant advances as Latin Americanists begin to take governmental institutional design more seriously than previously and to delineate the ways in which democratic institutions have combined in enduring ways. By focusing on national political institutions, they have underscored the different types of democracy that politicians can construct. These newly constructed institutions matter not least because they provide the rules and regulations that seek to order political interaction and make politics more transparent and predictable. Governmental institutions matter, therefore, for the formal locus and direction of political interaction.

But the current institutional trend in democratic consolidation studies is conceptually and analytically too sanguine. Conceptually, the literature defines democratic consolidation in narrow, dichotomous, and largely teleological terms that misrepresent an empirical context that is much more open-ended and nuanced.[60] The concept is narrow insofar as democratic consolidation is understood as the *absence* of regime breakdown following two consecutive and democratically held elec-

[59] There are scores of volumes and articles that explore this theme. See, in particular, Juan J. Linz, "The Perils of Presidentialism," *Journal of Democracy* 1 (Winter 1990); Matthew Soberg Shugart and John M. Carey, *Presidents and Assemblies: Constitutional Design and Electoral Dynamics* (Cambridge: Cambridge University Press, 1992); Juan J. Linz and Arturo Valenzuela, eds., *The Failure of Presidential Democracy: The Case of Latin America*, vol. 2 (Baltimore: Johns Hopkins University Press, 1994); Arend Lijphart and Carlos H. Waisman, eds., *Institutional Design in New Democracies: Eastern Europe and Latin America* (Boulder, Colo.: Westview Press, 1996); Scott Mainwaring and Matthew Soberg Shugart, eds., *Presidentialism and Democracy in Latin America* (Cambridge: Cambridge University Press, 1997); and Alfred Stepan and Cindy Skach, "Constitutional Frameworks and Democratic Consolidation: Parliamentarism versus Presidentialism," *World Politics* 46 (October 1993).

[60] Similar arguments are made by Guillermo O'Donnell, "Illusions about Democratic Consolidation," *Journal of Democracy* 7 (April 1996); idem, "Horizontal Accountability in New Democracies," *Journal of Democracy* 9 (July 1998); Kurt von Mettenheim and James Malloy, "Introduction," in Mettenheim and Malloy, eds., *Deepening Democracy in Latin America* (Pittsburgh, Pa.: University of Pittsburgh Press, 1998); Ben Ross Schneider, "Democratic Consolidations: Some Broad Comparisons and Sweeping Arguments," *Latin American Research Review* 30, no. 2 (1995), 216, 219, 220, 231; David Becker, "Latin America: Beyond Democratic Consolidation," *Journal of Democracy* 20 (April 1999), 139.

tions. However, no particular characteristics that define or identify democratic consolidation are actually delineated. This places scholars of democratic consolidation on shaky ground, since they can only identify the dependent variable by the *failure* of the given regime to function; this borders on conceptual circularity. Moreover, this concept leaves no room for ambiguity, contradiction, and contestation—as in cases where elections are held consecutively but where democratic norms, institutions, and practices are not in operation or in cases where national reforms intended to promote liberal democracy (such as decentralization) shelter illiberal and authoritarian local enclaves. Indeed, the concept's dichotomous and teleological characterization of consolidation belies a significant degree of social and political flux (evidenced by the postliberal challenge) that does not fit into this dual way of thinking about democracy or the conditions under which it is likely to endure. The concept assumes a kind of fixity and homogeneity that begs the broader questions of how institutions function, how social actors adapt to these institutions, and which conditions encourage (newly) mobilized actors to support, sidestep, and/or subvert democratic institutions. It assumes a political endpoint, much as an earlier literature on political development assumed a political terminus.[61] Yet when analyzing the Latin American cases comparatively and on the ground, one is struck by how much the term "consolidation" obscures the multiple and often contradictory processes that are occurring in the region's new democracies—some of which deepen and others of which undermine democratic norms, practices, and institutions. O'Donnell's work on low-intensity citizenship and on the absence of horizontal accountability in the new democracies raises similar concerns.[62]

This problem of conceptualizing the dependent variable, democratic consolidation, is exacerbated by the analytical privileging of certain independent variables over others. The democratic consolidation literature maintains a generally singular focus on a *narrow* set of political institutions—governmental and electoral institutions. By analyzing governmental institutions alone, these studies unsurprisingly bias their observation of the dependent variable and miss or gloss over the ways in which other independent variables, such as *states* and *social forces*, can impinge on the capacity to consolidate different aspects of democracy.

[61] I thank Atul Kohli for highlighting how this plea to move away from consolidation as an analytical concept parallels Huntington's classic argument about the need to move away from teleological and homogenizing ways of conceptualizing and analyzing political development. Samuel P. Huntington, "The Change to Change," in Roy C. Macridis and Bernard E. Brown, eds., *Comparative Politics: Notes and Readings* (Chicago: Dorsey Press, 1986).

[62] See O'Donnell (fnn. 1 and 60, 1998).

This narrow analytic focus on governmental institutions rests on a unstated but pervasive assumption that variation in state capacity ar legitimacy does not matter for regime outcomes. It assumes that stat are homogenized, capable, and national institutions and, as such, ca and do provide the political infrastructure in which democratic institu tions can function. As a result these scholars focus their attention c the presence and performance of governmental institutions in majc cities alone, and their work consequently projects a national descriptiv picture of democratic practices that is in fact only institutional an urban in scope. By definition, this kind of comparative analysis of gov ernmental institutions glosses over the unevenly institutionalized an contested terms of the Latin American *states* that are presumed to wor in the service of these new government institutions. It sidesteps the de gree to which state institutions (as independent or intervening vari ables) differentially shape, regulate, and/or monitor social behavic outside of urban centers, as demonstrated by circumstances giving ris to the postliberal challenge. Once one acknowledges the uneven reac and contested terms of the state, however, one is obliged at least to con sider how regimes can possibly "consolidate" the political institution practices, and norms of democracy in the absence of capable states. Fc as Linz and Stepan have argued in their more sociologically nuance analysis of democratic consolidation, the unevenness and frailty of ac tual states cannot bode well for the future of democracy.[63]

The increasingly narrow analytic focus of the democratic consolida tion literature on governmental institutions has other drawbacks a well. It slights analysis of the ways in which these new institutions in teract with, engender, and/or constrain (emerging) social forces. I tends to assume the relevant actors and how they will respond.[64] So to

[63] In contrast to the great majority of consolidation studies, Linz and Stepan (fn. 57) do questic how varied types of states might consolidate or undermine democracy. What I call the weak reach the state they refer to as the problem of usable bureaucracies, neither of which bodes well for dem cratic consolidation. They also contend, among other things, that democracies cannot consolida where there is a "stateness" problem, which they take to mean disputes over international (rather th internal) state boundaries and national membership (who can be a citizen rather than what citizensh might entail). According to this definition, they argue, there is no stateness problem in Latin Ameri (p. 16). Indeed, these authors conclude this because they assume away the national question in Lat America—one they find very prevalent in their other cases. Yet while national conflicts in Eastern E rope have emerged to make states and (presumed) nations coincide, in Latin America they ha emerged to force states to recognize the multiethnic diversity of its citizens, as we have seen.

[64] A notable few of these institutional studies (i.e., Mainwaring and Shugart, fn. 59) do analyze s cial support by established political actors and urban social groups. They are concerned with the wa in which institutional arrangements enable politicians to elicit support, build coalitions, and wres with policy questions without centralizing power in the hands of the executive and without engend ing legislative paralysis. But these studies overwhelmingly neglect to analyze whether and how th

it tends to assume preferences.[65] However, it would be hard to make the case that government institutions determine the relevant actors, define the preferences, and command social allegiance exactly as intended. Indeed, this article has highlighted how state reforms can unintentionally politicize new cleavages and mobilize new social actors who may support or resist new democratic institutions. Latin America's increasingly vocal indigenous movements and the postliberal challenge are particularly important examples of the contested terms of Latin America's new democracies.[66] The democratic consolidation literature, however, has been impervious to these developments, precisely because it assumes which actors are important, what their preferences are, and how they will behave, rather than also problematizing the (new) actors and preferences that have emerged. In assuming away ethnic cleavages, these studies have missed or perhaps underestimated the new ethnic movements that are challenging the very state institutions that underpin the presumed liberal foundations of contemporary democracy. The failure to analyze states as they interact with (new) social forces is unfortunate, for it is difficult to explain which democratic institutions can function and endure without analyzing the degree to which these institutions (and the state) determine political behavior and/or command organized

institutions incorporate and sustain social support beyond traditional forces and beyond capital areas areas—focusing as they do so closely on the institutions that are constructed and so little on whether these institutions will be maintained or disrupted by social actors and groups outside of the state.

[65] Ames (fn. 58) also makes this point (p. 234).

[66] Black movements, women's movements, and poor people's movements, where they exist, have also demanded equal inclusion and greater access to state resources in Latin America's democratic regimes. Their emergence in Latin America highlights the failure of democratic institutions in many instances to incorporate social sectors that have historically been marginalized. Unlike the indigenous movements discussed in this article, however, their demands in most cases do not necessarily challenge the assumptions and terms of liberal democratic institutions and state formation as much as demand an equal footing in the regime. It is also important to note that these movements, in contrast to the region's indigenous movements, have declined in strength during the contemporary democratic regimes as political parties displace and/or absorb them on the political stage. On black movements, see fn. 4. On women's movements, see Sonia E. Alvarez, *Engendering Democracy in Brazil: Women's Movements in Transition Politics* (Princeton: Princeton University Press, 1990); Jane S. Jaquette, ed., *The Women's Movement in Latin America: Participation and Democracy*, 2d ed. (Boulder, Colo.: Westview Press, 1994); Amy Conger Lind, "Power, Gender, and Development: Popular Women's Organizations and the Politics of Needs in Ecuador," in Arturo Escobar and Sonia E. Alvarez, eds., *The Making of Social Movements in Latin America: Identity, Strategy, and Democracy* (Boulder, Colo.: Westview Press, 1992); and Jane S. Jaquette and Sharon L. Wolchik, eds., *Women and Democracy: Latin America and Central and Eastern Europe* (Baltimore: Johns Hopkins University Press, 1998). On poor people's movements, see Robert Gay, *Popular Organization and Democracy in Rio de Janeiro: A Tale of Two Favelas* (Philadelphia: Temple University Press, 1994); Philip D. Oxhorn, *Organizing Civil Society: The Popular Sectors and the Struggle for Democracy in Chile* (University Park, Pa.: Penn State University Press, 1995); Cathy Lisa Schneider, *Shantytown Protest in Pinochet's Chile* (Philadelphia: Temple University Press, 1995); and Susan S. Stokes, *Cultures in Conflict: Social Movements and the State in Peru* (Berkeley: University of California Press, 1995).

social support/apathy/opposition.[67] Once we integrate state and societ back into the contemporary picture, we are reminded that governmental institutions do not operate exactly as planned, nor do they necessarily operate similarly across cases.[68]

This conceptual and analytical discussion suggests that the focus o "democratic consolidation" is misplaced—particularly in new democracies.[69] Rather, we should develop a new conceptual research agenda tha replaces this teleological dependent variable with more conceptuall and analytically nuanced studies of *democratic politics*. The dependen variable would identify the varied *scope*, *depth*, and *blend* of democrati institutional reforms, political norms, and practices.[70] The task woul be to explain the variation in the existing outcomes, rather than assuming a golden democratic endpoint that regimes have or have no reached (an assumption that many contemporary consolidation studie characteristically tend to make). This more broadly conceived analysi of democratic politics would generate a dynamic and consequential se of research questions about the uneven and often contradictor

[67] Indeed, the failure to problematize how emerging social forces (particularly in the countryside contest the process of institutionalizing democracy weakens the analytic narrative and causal argu ments in democratic consolidation studies. This is because regime endurance (the dependent variabl in consolidation studies) is fundamentally linked to the politics of the countryside. A substantial com parative historical literature has argued and illustrated how regime endurance requires states to secur control, if not command loyalty, of the countryside. These studies have taken institutions seriously, bu they have also problematized the process of building institutions and of cultivating the social force that could support those institutions. The failure of contemporary democratic consolidation studies t incorporate this central insight curtails the power of their arguments about the links between institutions and regime endurance. See Barrington Moore, Jr., *Social Origins of Dictatorship and Democracy Lord and Peasant in the Making of the Modern World* (Boston: Beacon Press, 1966); Samuel P. Huntington, *Political Order in Changing Societies* (New Haven: Yale University Press, 1986); Theda Skocpol *States and Social Revolutions: A Comparative Analysis of France, Russia, and China* (Cambridge: Cambridge University Press, 1979); Collier and Collier (fn. 8); Timothy R. Scully, *Rethinking the Center Party Politics in Nineteenth- and Twentieth-Century Chile* (Stanford, Calif.: Stanford University Press, 1992); Evelyne Huber and Frank Safford, eds., *Agrarian Structure and Political Power: Landlord an Peasant in the Making of Latin America* (Pittsburgh, Pa.: University of Pittsburgh Press, 1995); Deborah J. Yashar, *Demanding Democracy: Reform and Reaction in Costa Rica and Guatemala, 1870s–1950* (Stanford, Calif.: Stanford University Press, 1997); Jonathan Fox, ed., The Challenge of Rural Democratisation: Perspectives from Latin American and the Philippines, *Journal of Development Studies*, special issue, 26 (July 1990); and Edward L. Gibson, "The Populist Road to Market Reform: Policy and Electoral Coalitions in Mexico and Argentina," *World Politics* 49 (April 1997).

[68] It is striking that the democratic consolidation literature pays no heed to an earlier institutional literature on democratic stability in divided societies—a literature that concluded that there is no one model for all societies.

[69] For parallel arguments, see fn. 60.

[70] See Philippe C. Schmitter, "The Consolidation of Democracy and Representation of Social Groups," *American Behavioral Scientist* 35 (March–June 1993). Schmitter states (p. 444): "The label *democracy* hides a continuous evolution in rules and practices and an extraordinary diversity of institutions." While Schmitter seems unwilling to dispense with the concept of consolidation, he does call for a more disaggregated study of the new democracies—one that would analyze political democracy as a composite of partial and competing regimes. He also calls for studies of the *types* of democracies that have emerged and notes the importance of accounting for the causal role of emerging associations.

processes of democratic change. It would compel scholars to identify, analyze, and explain the diverse and at times contradictory patterns emerging in Latin America's new democracies in arenas as significant as citizenship regimes, the rule of law, decentralization, patterns of participation, and the terms of contentious politics.

If we need to reconceptualize the dependent variable to accommodate a much more dynamic and varied set of democratic outcomes, so too we need to evaluate a more expanded repertoire of independent variables in our research and our writing. Rather than assuming the a priori significance of government institutions, we should be required to compare the causal role of these institutions against other factors (including the state, social forces, and the economy) to ascertain if, when, and why a given variable is most important. In some cases governmental institutional design will be found to be the primary factor; in others, it will be found to be secondary or even inconsequential. This process of evaluating competing independent variables would be designed not to lead to more idiographic analyses but rather to hold scholars more accountable and to demand more rigor in their comparative explanations of which variables matter and why.

A research agenda of democratic politics would reopen an older set of questions that would include, but not be limited to, the following: How has the sequencing of democratic reforms affected their scope, depth, and institutionalization? Why have different countries experimented with different kinds of political reform (for example, decentralization) and what impact have they had on sectoral participation, accountability, and political accommodation? Why have (ethnic or other) cleavages been politicized in some new democracies but not in others? Why do states respond differently to multiethnic demands and how do we explain both the varied responses and the impact of the resulting political reforms? Why have some democratic reforms (for example, judicial reform and the institutionalization of the rule of law) been so elusive, even as others have been widely implemented? What impact, if any, has globalization had on democratic accountability and participation?

In short, rather than assuming that government institutions alone can consolidate democracy, scholars should pay closer attention to how the sequencing, pace, and institutionalization of democratic reforms actually affects the depth and scope of democratic politics. This conceptual shift in foci requires exchanging a teleological and singularly institutional approach for a more analytically nuanced one that analyzes the interplay between political change and contestation in state and so-

ciety. While this approach is perhaps more ambitious than a simple i stitutional approach that focuses on government institutions alone, i sure to yield more dynamic insights about the direction and terms politics in the region's new democracies. Otherwise stated, if we are understand the challenges for third-wave democracies—challenges th affect both democratic quality and endurance—we cannot search the institutional holy grail of democratic consolidation.

APPENDIX: ACRONYMS FOR INDIGENOUS AND/OR PEASANT ORGANIZATIONS

AIDESEP	Asociación Interétnica de Desarrollo de la Selva Peruana (Peruvian indigenous organization in the Amazon)
EZLN	Ejército Zapatista de Liberación Nacional (Mexican indigenous organization based in Chiapas)
CIDOB	Confederación Indígena del Oriente, Chaco y Amazonía de Bolivia (Bolivian regional indigenous organization in the lowlands)
CNC	Confederación Nacional Campesina (Mexican national peasant organization)
COICA	Coordinadora de Organizaciones Indígenas de la Cuenca Amazónica (Amazonian indigenous organization with participation by indigenous organizations whose countries are partially located in the Amazonian Bas
COMG	Consejo de Organizaciones Mayas de Guatemala (Guatemalan national indigenous organization)
CONAIE	Confederación de Nacionalidades Indígenas del Ecuador (Ecuadorian national indigenous organization)
CPIB	Central de Pueblos Indígenas del Beni (Bolivian Amazonian indigenous organization based in the Beni)
CSUTCB	Confederación Sindical Única de Trabajadores Campesinos de Bolivia (Bolivian national peasant organization that has incorporated indigenous demands)
OPIP	Organización de Pueblos Indígenas del Pastaza (Ecuadorian Amazonian indigenous organization based in Pastaza)

LOYALTY DILEMMAS AND MARKET REFORM
Party-Union Alliances under Stress in Mexico, Spain, and Venezuela

By KATRINA BURGESS*

DURING the twentieth century workers in Latin America and Western Europe were often incorporated into unions with institutionalized links to political parties. In the wake of economic crisis, globalization, and market reform, these links are coming under stress. Unions find themselves struggling to maintain their membership and to respond effectively to the double movement of globalization and decentralization, and they see their mediating capacity and political legitimacy being severely tested in many cases by the adoption of market reform by their party allies in government. The result is a crisis of representation that threatens to alter profoundly the mechanisms by which workers are (or are not) linked to the state.

This article analyzes the impact of these trends on historic alliances between governing parties and labor unions in Mexico, Spain, and Venezuela. All three countries were governed in the 1980s and early 1990s by parties with historic ties to the labor movement. In Mexico the Institutional Revolutionary Party (PRI) or one of its antecedents had been organically linked to the Confederation of Mexican Workers (CTM) since the 1930s. In Spain the Socialist Party (PSOE) had been closely associated with the General Workers' Union (UGT) since the 1880s. And in Venezuela, Democratic Action (AD) had dominated the Confederation of Venezuelan Workers (CTV) since the 1940s. These parties owed much of their success to the support of affiliated unions,

* I am grateful for comments at various stages of this project from Jeremy Adelman, Peter Andreas, Nancy Bermeo, Miguel Centeno, Stephan Haggard, Robert Kaufman, Kevin Middlebrook, John Waterbury, my colleagues in the Global Political Economy Research Consortium of the Global Affairs Institute in the Maxwell School of Citizenship and Public Affairs at Syracuse University, and three anonymous reviewers. Funding for my fieldwork in Mexico, Spain, and Venezuela was provided by the Council of European Studies at Columbia University and the Department of Politics, the Program in Latin American Studies, and the Center of International Studies at Princeton University.

which, in turn, enjoyed a degree of influence that outweighed their organizational presence among workers.

In the 1980s and early 1990s all three parties adopted economic reforms that violated their long-standing principles by imposing painful sacrifices on organized labor. These reforms—which included fiscal and wage austerity, trade liberalization, price deregulation, flexibilization of the labor market, privatization of state-owned enterprises, industrial restructuring, and welfare reform—subjected workers to declining incomes, unemployment, decreased job security, and cuts in social services. They also reduced the mediating capacity of unions by shrinking the size of the public sector and expanding the influence of neoliberal technocrats in policy-making circles. Yet party-affiliated labor leaders responded differently in each case. In Mexico they complained bitterly about the reforms but neither organized serious protest nor took steps to extricate themselves from the alliance. In Venezuela they came closer to breaking their ties with the party but ultimately retreated to their traditional strategy of negotiating behind closed doors. Only in Spain did they abandon their historic ties with the party in response to its market reforms.

I argue in this article that the strategic choices of disaffected labor leaders can be understood with reference to Albert Hirschman's schema of exit, voice, and loyalty.[1] I propose that voice and exit constitute points along a path of alliance change, and that loyalty interacts with authority structures to affect the pace and direction of this change. Although loyalty is a very underdeveloped concept in Hirschman's model, I find that it is critical to the fate of party-union alliances under stress. To understand how and why such alliances change, we need to know how loyalty is constructed, of what it consists, and under what conditions it is likely to evaporate. I seek to unravel to these puzzles by examining the role of loyalty in the fate of party-union alliances under stress in Mexico, Spain, and Venezuela.

Loyalties under Stress

Political parties and labor unions began to forge alliances in Latin America and Western Europe with the advent of industrialization and mass politics in the late nineteenth and early twentieth centuries.[2] Al-

[1] Hirschman, *Exit, Voice, and Loyalty* (Cambridge: Harvard University Press, 1970).

[2] Alessandro Pizzorno, "Political Exchange and Collective Identity in Industrial Conflict," in Colin Crouch and Alessandro Pizzorno, eds., *The Resurgence of Class Conflict in Western Europe since 1968* (New York: Holmes and Meier, 1978); Ruth Berins Collier and David Collier, *Shaping the Political Arena* (Princeton: Princeton University Press, 1991).

though the earliest unions tended to be apolitical or anarchist, unions and parties discovered shared interests as workers developed a class consciousness, the state increased its involvement in the economy, and working-class support became a valuable commodity in the struggle for state power. These conditions laid the foundation for an exchange whereby the union would help the party win or maintain control of the state in return for the party's commitment to use state intervention to channel benefits to labor. Although many labor-backed parties initially advocated radical state intervention in the form of socialism or communism, participation in the political system usually led them to moderate their programs in favor of using state power to promote the interests of workers in the context of a capitalist economy.[3]

These exchanges facilitated the construction of stable, institutionalized alliances between parties and unions that were informed by strong bonds of loyalty. Rather than being ad hoc and dictated by short-term interests, these alliances resembled "regimes" as understood by international relations scholars. Stephen Krasner defines regimes as "sets of principles, norms, rules, and decision-making procedures around which actors' expectations converge."[4] Although he and his colleagues apply the concept of regimes to nation-states in the international arena, it can be extended quite naturally to organizations in the domestic arena. A defining feature in both cases is the nonunitary character of the central actors. Like nation-states, parties and unions have complex internal structures, identities, and constituencies. They can enter into regularized patterns of interaction with one another, but they do so as collective entities rather than as individuals. Thus, they must consider more than their individual preferences; they must also take into account the potential responses of their followers. In addition, these central actors are relatively few in number, which means any one of them can have an impact on institutional outcomes.

Like regimes, party-union alliances are structured by principles and norms.[5] The principles elaborate their shared vision of society; and the norms (1) lay out each partner's side of the bargain in the socioeconomic, political, and organizational arenas and (2) govern the management of negotiation and conflict within the alliance. Whereas some

[3] Gøsta Esping-Andersen, *Politics against Markets* (Princeton: Princeton University Press, 1985): Adam Przeworski and John Sprague, *Paper Stones* (Chicago: University of Chicago Press, 1986).

[4] Krasner, "Regimes and the Limits of Realism," in Stephen D. Krasner, ed., *International Regimes* (Ithaca, N.Y.: Cornell University Press, 1983), 2.

[5] Although rules and procedures are also important, I argue that they can be subsumed under norms.

principles and norms will be explicit and formalized, others will consist of informal yet mutual understandings that have evolved out of repeated interactions between the party and its labor allies. Both principles and norms serve as codes of conduct that alter the expectations and, hence, the behavior of alliance participants. In particular, these participants become more inclined to trust their exchange partners to behave in ways that are favorable to their goals and interests.

Most party-union alliances in Latin America and Western Europe in the twentieth century developed around the principle of state intervention in the economy. This principle generated similar norms regarding the rights and obligations of each partner. In the socioeconomic arena the party was expected to use the power of the state, when available, to deliver wage and nonwage benefits to labor, as well as to adopt procedures that gave labor leaders a role in formulating and implementing relevant policies. In return affiliated unions were expected to moderate their strike activity and to refrain from maximizing their wage demands when the party was in government, particularly in times of economic crisis. In the political arena the party was expected to give labor leaders access to public office and party posts in return for the union's delivery of the working-class vote in elections. Finally, in the organizational arena the party was expected to seek (or maintain) the hegemony of its labor allies in the labor movement in return for legitimation as a representative of the popular classes. Although the party could deliver many of these benefits only when in government, it exchanged the *promise* of their delivery for the union's support when in the opposition.

Over time these principles and norms encouraged longer time horizons and a general sense of obligation toward the alliance. In other words, they reinforced bonds of loyalty between the party and its labor allies. As suggested by several scholars, loyalty can take the form of either an attitude or a behavior.[6] A loyal attitude can be defined as a belief in the good intentions of another actor. Even if the actor strays in the short term, the loyalist believes he or she will come around in the long term. This belief may derive from ideological affinities, personal relationships, and/or past responsiveness to the loyalist's demands.

[6] See, e.g., Morton Grodzins, *The Loyal and the Disloyal* (Chicago: University of Chicago Press, 1956), 5; Harold Guetzkow, *Multiple Loyalties: Theoretical Approach to a Problem in International Organization* (Princeton: Center for Research on World Political Institutions, Princeton University, 1955), 8; Yossi Shain, *The Frontier of Loyalty* (Middletown, Conn.: Wesleyan Unversity Press, 1989), 22; and Michael J. Withey and William H. Cooper, "Predicting Exit, Voice, Loyalty, and Neglect," *Administrative Science Quarterly* 34, no. 4 (1989), 522.

ınce loyalty, like trust, grows with use and decays with disuse, reserves
loyalty can be accumulated and then tapped in times of stress.[7] But
ıese reserves must ultimately be replenished through actions that reaf-
m the loyalist's faith in the other actor's good intentions. Rather than
ing blind, a loyal attitude rests on rational expectations about the fu-
re and is likely to be replaced by feelings of betrayal if these expecta-
ons are dashed.

Loyal behavior can be defined as living up to one's explicit or implict
mmitments to another actor. In the case of a party-union alliance, a
yalist conforms to the principles and norms on which the alliance is
ısed. Although a loyal attitude and loyal behavior are mutually rein-
rcing, the latter can exist independently of the former. As John
haar notes, "Active disloyaly is very difficult: custom, the climate of
inion, informal and formal sanctions, inertia, fear, the lack of clear
ternative objects of loyalty—all these forces work to assure that even
ose who are not actively and intensely loyal are at least not *dis*loyal."[8]
hus, loyal behavior may persist even after the loyalist loses faith in an-
her actor's good intentions. Once the reserves of loyal attitude have
en depleted, however, it becomes more difficult to prevent outbreaks
disloyal behavior.

Building reserves of loyalty between parties and unions was relatively
sy in the context of mass production, limited global competition, and
ced exchange rates. First, labor-backed parties could make credible
omises to use state power to deliver concrete benefits to unions and
orkers. And at least until the 1980s they usually fulfilled these
omises when in government. In Western Europe they tended to pur-
e Keynesian demand management in the context of a mixed-
onomy welfare state. In Latin America they tended to adopt a de-
and-driven model of import-substitution industrialization (ISI) com-
ned with limited welfare benefits for workers and the middle classes.
oth strategies involved extensive state intervention in the economy
ıd granted affiliated unions privileged access to economic, political,
ıd organizational benefits. Although labor-backed parties sometimes
treated from their commitments in times of economic crisis, these

[7] Partha Dasgupta, "Trust as a Commodity," in Diego Gambetta, ed., *Trust: Making and Breaking
operative Relations* (Oxford: Basil Blackwell, 1988), 56. For a related and intriguing discussion of
w trust mediates the perceptions and responses of employees regarding the breach of psychological
ntracts by their employers, see Sandra L. Robinson, "Trust and Breach of the Psychological Con-
ct," *Administrative Science Quarterly* 41, no. 4 (1996).

[8] Schaar, "Loyalty," in *International Encyclopedia of the Social Sciences*, vol. 9 (New York: Macmillan
mpany and the Free Press, 1968), 485.

setbacks were usually only temporary, thereby reinforcing the loyalty affiliated unions.

Second, affiliated unions were well positioned to deliver work support and cooperation to the party. As a relatively homogeneoı group in stable work roles, workers tended to identify strongly wi one another and with the union claiming to represent them. In add tion, they were often geographically concentrated, which reinforc class identities and made it easier for the union to monitor complian with economic and political directives. Workers also relied quite hea ily on the union for economic and political information. According Victor Pérez Díaz, these conditions facilitated a relationship of "in plicit delegation" between workers and unions, meaning that worke trusted the unions to look out for their interests and did what tl unions asked of them.[9] These strong and stable loyalties helped tl unions deliver on their promises to the party, even during times of ec nomic crisis.

In this context party-affiliated unions had the opportunity to med ate between the party and workers, particularly when the party was government. In effect, they became managers of a virtuous circle of lo alty. Party leaders reinforced the loyalty of labor leaders by offerir them policies favorable to workers and unions, union subsidies, acce to public office, and participation in policy formulation. Labor leade in turn, used these resources to reinforce the loyalty of workers by o fering them higher wages, increased employment, job security, welfa benefits, and other socioeconomic improvements. The loyalty of the workers was then translated by labor leaders into electoral and organ zational support for the party, thereby reinforcing the party's loyalty labor. Although these loyalties were never entirely unconditional, rel tively stable identities and predictable economic trends allowed for tl accumulation of reserves of loyalty that could be tapped in times stress and replenished in times of prosperity.

These conditions favoring the building of reserves of loyalty betwe parties and unions began to deteriorate in the late twentieth centu however, as both identities and economic conditions became mo fluid. Heightened international competition, the collapse of the Bre ton Woods system of fixed exchange rates, the information revolutio and the growth of the service sector all contributed to more flexib forms of economic organization in the 1970s and 1980s. In the proce

[9] Pérez-Díaz, "Unions' Uncertainties and Workers' Ambivalence," *International Journal of Politi Economy* 17 (Fall 1987), 120.

'orker identities fragmented and multiplied while the traditional base f union support declined as a share of the workforce. These changes ıade it "increasingly difficult for trade unions to secure workers' loy-lty on the basis of class ideology and collective identity."[10] As a result, ›yalty reserves became more vulnerable to depletion in the event that ıe union stopped delivering benefits and protection to individual 'orkers in the short term.

Unfortunately, changes in the global economy have reduced the ca-acity and willingness of labor-backed parties to facilitate this delivery 'hile in government. First, international competition and the relative ecline of mass production have undermined both the rationale and the :sources for state intervention in the economy. Second, the same un-:rtainties that are leading employers to seek more flexible forms of roduction are encouraging parties to pursue more flexible strategies of ɔalition building and policy-making. Thus, many labor-backed parties ave begun to diversify their support base away from organized labor, › shift policy-making authority from labor leaders to neoliberal tech-ocrats, and to cultivate direct links with individual voters, thereby by-assing the unions as intermediary organizations.

RESPONSES BY DISAFFECTED LABOR LEADERS

Vhen labor-backed parties reduce state intervention and/or engage in ɔalitional restructuring, they violate alliance principles and norms and reak the virtuous circle of loyalty that has sustained their ties with ıbor. In the process they devalue the terms of labor's exchange with the arty. But, as we see in Mexico, Spain, and Venezuela, not all labor :aders respond to this devaluation in the same way. Borrowing from Iirschman's seminal work, *Exit, Voice, and Loyalty*, I argue that they 'ill choose one of three responses: norm-based voice, norm-breaking oice, or exit. Their choice depends on how loyalty interacts with the uthority structures in which labor and party leaders find themselves. lesides determining their relative vulnerability to reprisals for disloyal ehavior, these authority structures affect the party's capacity to sustain ıeir loyalty either by offering them forms of compensation consistent 'ith the reforms or by taking their side against the reforms.

[10] Mario Regini, "Industrial Relations in the Phase of 'Flexibility,'" *International Journal of Political conomy* 17 (Fall 1987), 92.

EXIT AND VOICE IN PARTY-UNION RELATIONS

Hirschman posits exit and voice as alternative responses by custom of a firm or members of an organization to a decline in the quality the goods and services they are receiving. Exit occurs when they aba don the relationship altogether. Voice occurs when they make "any a tempt to change, rather than to escape from, an objectionable state affairs."[11] Hirschman finds that exit is the most likely response becau it is relatively costless and straightforward. Once an individual h gathered information about alternative products and organizations, l or she can exit without investing in political resources or organizin collectively. Voice, by contrast, "is costly and conditioned on the influ ence and bargaining power customers and members bring to bea within the firm from which they buy or the organization to which the belong."[12]

Party-affiliated labor leaders face an analogous set of choices whe the party adopts market reform and thereby devalues their terms of ex change. Assuming they are unwilling to remain silent, they can eithe pressure the party to reverse the decline (voice) or defect from the al liance altogether (exit). Unlike the atomized individuals in Hirschman model, however, they are likely to find voice to be less costly than exi Exit is discouraged by the uncertainty and risk associated with aban doning regularized patterns of interaction.[13] Meanwhile, voice is facili tated by the channels of communication already built into the allianc As argued above, norms will have emerged to govern the managemen of negotiation and conflict among alliance participants. Conforming t these norms is likely to be much less costly than violating them, an uncertainty surrounding the application of voice is likely to be lowere by previous episodes of responsiveness by the party.

In addition, the small number of actors in a party-union alliance i likely to raise both the resonance of voice and the costs of exit. In part this is true because exit by a single actor (for example, the union) wil lead to a collapse of the alliance, whereas organizations or firms ca usually withstand the defection of numerous members or customer Assuming the party still wants to preserve the alliance, this possibilit lends greater weight to the threat of exit, thereby enhancing the effec tiveness of voice. But by the same token, exit itself cannot be under taken quietly or inconspicuously. It is inevitably a loud, political act

[11] Hirschman (fn. 1), 30.

[12] Ibid., 40.

[13] Arthur L. Stinchcombe, *Constructing Social Theories* (New York: Harcourt, Brace and Worl 1968), 123.

TABLE 1
LABOR LEADER RESPONSES TO MARKET REFORM

	Norm-Based Voice	Norm-Breaking Voice	Exit
Mexico (CTM leaders)	1982–94		
Venezuela (CTV leaders)	1984–89 →	1989–90	
	1990–93 ←		
Spain (UGT leaders)	1982–84 →	1985–88 →	1988–93

Combined with the other obstacles to exit, these factors are likely to nake voice the default option for disaffected labor leaders in a party-union alliance.

We have not yet exhausted the menu of options facing these leaders, nowever. Because of alliance norms, they have the possibility of more han one kind of voice. The default option, which I call *norm-based voice*, involves tactics that conform to the norms that have emerged to govern the management of negotiation and conflict in the alliance. But f these tactics fail to elicit the desired response, disaffected labor leaders can choose a second option that falls short of exit. This option, which I call *norm-breaking voice*, violates alliance norms but does not constitute abandonment of the relationship. In fact, it is likely to coexist with norm-based voice on other issues or at other times.[14] Thus, while risky, norm-breaking voice does not dispense entirely with the egularized patterns of interaction that have historically served these eaders so well. It is only when this kind of voice falls on deaf ears that abor leaders are likely to resort to exit.

I argue that a shift by disaffected labor leaders from one option to another constitutes movement along a continuum from norm-based voice to exit. As illustrated in Table 1, labor leaders in Mexico, Spain, and Venezuela moved to different points along this continuum during he reform period.[15] In Mexico, CTM leaders complained bitterly about

[14] Although the line between norm-based voice and norm-breaking voice depends on the norms hat have emerged in each party-union alliance, a common example is the distinction between issuing *threats* to hold a general strike (norm-based voice) and actually mobilizing a general strike (norm-breaking voice) when the party is in government.

[15] For a detailed discussion of the response of affiliated unions to the party's economic reforms in each case, see Katrina Burgess, "Alliances under Stress: Economic Reform and Party-Union Relations in Mexico, Spain, and Venezuela" (Ph.D. diss., Princeton University, 1998).

the economic reforms and threatened mobilizations to improve their negotiating position with the PRI, but they neither organized serious protest against the reforms nor took steps to extricate themselves from the party-union alliance. In fact, they collaborated with industrial restructuring in return for control over collective contracts, signed a series of anti-inflation pacts with business and the government, supported the negotiation of a free-trade agreement with the United States and Canada, fought hard to preserve their institutionalized presence within the PRI, and actively campaigned for PRI candidates in elections. Rather than risking their ties with the governing party, they chose to remain within the confines of norm-based voice.

Venezuela's labor leaders came closer to breaking their alliance with the governing party. After several years of limiting their voice to rhetorical attacks on the economic cabinet and threats to mobilize workers against the government's austerity policies, the CTV joined with Venezuela's other labor confederations in May 1989 to call a general strike against the reforms of the new AD government of Carlos Andrés Pérez. The strike was an unprecedented act of norm-breaking voice that brought the economy to a halt and severely damaged the government's credibility. Venezuela had not experienced a general strike since 1958, when AD and its labor allies had joined with other prodemocracy forces to protest an attempted military coup. Nor had the CTV ever taken such a combative stance while AD controlled the presidency. But rather than representing a step down the path toward exit, the general strike turned out to be an exception to the rule of party-union collaboration. Despite the failure of the general strike to reverse the reforms, labor leaders quickly returned to a strategy of norm-based voice.[16]

Only in Spain did affiliated labor leaders exit from their historic alliance with the governing party. As in Mexico and Venezuela, they initially responded to market reform with norm-based voice. Between 1982 and 1984 they cooperated reluctantly with the PSOE's policies of fiscal austerity, industrial restructuring, and labor market reform. They soon moved into norm-breaking voice, however, by breaking with party discipline in the parliament to protest a pension reform in 1985, resigning their parliamentary seats rather than vote for the PSOE's budget in 1987, and, finally, joining with the communist-affiliated union, the Workers' Commissions (CCOO), to hold a general strike against the

[16] An eloquent example of this return to norm-based voice was the decision by AD's labor delegation to vote with the party to reject a motion of censure against Pérez's cabinet in March 1992, despite the CTV's strong criticism of the cabinet and its policies. *El Nacional*, March 27, 1992.

government's economic policies in December 1988.[17] Besides representing an all-out assault on the government's credibility, the strike was of tremendous symbolic importance. Spain had previously experienced only four truly general strikes, all of which took place at moments of regime crisis. And whereas the PSOE and the UGT had joined forces in the earlier strikes, they were on opposite sides of the picket line in 1988.

In contrast to Venezuela, the general strike in Spain carried the UGT over the threshold from norm-breaking voice to exit.[18] Over the next few years the UGT and the PSOE severed the remaining symbolic and institutional ties on which their century-long relationship had been based. In the legislative elections of 1989 the UGT failed to back the PSOE for the first time in its history, and its top leader, Nicolás Redondo, publicly hinted at his support for the opposition coalition, the United Left. UGT leaders also refused to negotiate global economic pacts with the socialist government and joined with the CCOO to hold two more general strikes, in May 1992 and January 1994. Meanwhile, the PSOE launched a project to create a network of militants in the factories to fulfill the UGT's earlier role of disseminating information and mobilizing support. In November 1990 the delegates at the PSOE's Thirty-second Congress voted to eliminate obligatory affiliation with the UGT so as to allow labor leaders affiliated with other unions, particularly the CCOO, to serve on the party's governing bodies.

These contrasting stories of party-union relations raise two questions. First, why did labor leaders in all three countries respond initially with the same strategy of norm-based voice? Second, why did their strategies eventually diverge? I argue that both questions can be answered with reference to loyalty. Labor leaders responded initially with norm-based voice because they were all drawing on the reserves of loyalty that the party had accumulated in previous interactions with them. But their strategies diverged over time because they faced different opportunity costs to remaining loyal to the party as it adopted market reform.

[17] Although the catalyst for the 1988 general strike was a Youth Employment Plan to allow temporary contracts for workers under twenty-five years of age, the UGT demanded a fundamental reorientation in the government's economic program. For a thorough discussion of the UGT's demands, see the articles in *Claridad* 29–30 (1989).

[18] Exit did not mean that UGT leaders stopped negotiating with the PSOE, however. Since workers and unions could not avoid the consequences of the party's policies even after the collapse of the alliance, they continued to exercise "voice after exit." See Hirschman (fn. 1), 101, 104.

LOYALTY AND LOYALTY DILEMMAS

I have already suggested how loyal attitude and loyal behavior are embedded in a party-union alliance. With regard to loyal attitude, historically constructed expectations regarding the good intentions of th party produce reserves of loyalty among labor leaders. As long as thei expectations are kept alive—and the loyalty reserves are thereby maintained—labor leaders are likely to choose voice over exit in response t a deterioration in their terms of exchange.

Hirschman makes a similar argument that loyalty "holds exit at ba and activates voice."[19] Unfortunately, he never defines loyalty, whicl leaves him open to the criticism that he is using it as "an *ad hoc* equatio filler."[20] According to Brian Barry, Hirschman predicts the probabilit of exit on the basis of (1) a cost-benefit analysis concerning the value o the best alternative and the probability of gaining an improvement i the existing product or organization at a given cost of voice; and (2 loyalty, which acts as a tax on competing products or organizations. Bu because he provides no independent criteria by which to measure loyalty, "the equation can always be made to fit the facts *ex post,* by imputing loyalty in sufficient quantity to a person who, on the basis of th first term (the cost-benefit calculation), should have switched but i fact has not done so."[21]

We can overcome this problem by defining loyalty as historicall constructed expectations regarding the good intentions of anothe actor. Rather than constituting an ad hoc tax on competing product or alternatives, loyalty becomes part of the first term, that is, the probability that voice will produce results. Hirschman in fact suggests thi solution when he argues that loyalty involves a "reasoned calculation based on previous interactions that voice will produce results.[22] He late builds on this idea by linking loyalty to trust.[23] As long as disaffecte

[19] Hirschman (fn. 1), 78. Note that loyalty is not equivalent to silence. A decision to remain with product or organization does not necessarily lead to the exercise of voice. Instead, an actor may g along with a deteriorating status quo without engaging in pressure tactics to reverse it. Brian Barr "Review Article: Exit, Voice, and Loyalty," *British Journal of Political Science* 4, no. 1 (1974), 95. Bu rather than expanding the menu of options to include "silent non-exit," many scholars mistakenly us loyalty in its stead. This move creates a dichotomy between loyalty and voice that is inconsistent witl Hirschman's model.

[20] Barry (fn. 19), 95.

[21] Ibid., 96.

[22] Hirschman (fn. 1), 79, 91. Barry recognizes this solution when he writes that "[w]hat I thinl [Hirschman] wants to say here is that the experience of having had influence in a collectivity and th expectation of having influence in the future tend to increase a person's commitment to that collectivity, in other words loyalty to it in the ordinary meaning of 'loyalty.'" Barry (fn. 19), 99.

[23] Albert O. Hirschman, *Rival Views of Market Society and Other Recent Essays* (New York: Viking 1986), 81.

TABLE 2
SOLUTIONS TO THE LOYALTY DILEMMAS

	Disloyalty		
	Solution 1	*Solution 2*	*Solution 3*
ırty leaders toward executive	no	no	yes
abor leaders toward workers	yes	no	no
abor leaders toward party	no	yes	no

ıbor leaders trust their party allies to honor their demands, they are kely to continue relying on voice. If that trust is betrayed, however, ıey are likely to cross the threshold into exit unless other inhibiting ıctors are present.

With regard to loyal behavior, alliance principles and norms create :laims" on the loyalty of each alliance participant. If disaffected labor ·aders limit their resistance to norm-based voice, they are respecting ıe loyalty claims of the party. Once they move into norm-breaking ›ice, they begin violating these claims and thereby behave disloyally ›ward the party. Crossing the threshold into exit constitutes the ultiıate act of disloyalty because it brings about the collapse of the alance.[24] But the party's claims are not the only ones that are relevant. ecause parties and unions are collective entities rather than unitary ac›rs, their leaders face multiple claims on their loyalty and may somemes find themselves pulled in strategically contradictory directions.

This is exactly what happens when a labor-backed party in governıent adopts reforms that impose hardships on workers and unions. ·aught in a "loyalty dilemma," labor and party leaders are left with no ıoice but to behave disloyally toward one set of claimants. For labor aders, the dilemma involves a choice between the loyalty claims of the ırty and the loyalty claims of their followers. Either they tolerate the :forms at the risk of losing worker support, or they resist the reforms · the risk of losing access to resources controlled by the party.[25] For ırty leaders, the dilemma involves a choice between the loyalty claims : their own executive and the loyalty claims of their labor allies. Either ıey close ranks with the executive at the risk of losing the support of

[24] As suggested above, disaffected labor leaders may continue to adhere to alliance norms even after :ir reserves of loyalty toward the party have been depleted, especially if the party has the capacity to nish them for disloyalty. At this point, however, loyal behavior is no longer motivated by loyal attile and thus becomes highly sensitive to short-term calculations of interest and power.

[25] For a similar discussion of labor leaders who are pulled in strategically contradictory directions by : policies of labor-backed parties in government, see Peter Gourevitch et al., *Unions and Economic isis: Britain, West Germany, and Sweden* (London: George Allen and Unwin, 1984), 11.

labor, or they join labor in opposition to the reforms at the risk of lo ing access to resources controlled by the executive.[26]

As illustrated in Table 2, these overlapping loyalty dilemmas ha three possible solutions, each of which has different implications f the fate of the alliance. In solution 1 labor leaders behave disloyally t ward workers. This solution is consistent with alliance survival becau labor leaders continue to honor their commitments to the party desp losing their capacity to channel significant inducements to worke The likely price of survival, however, is a heavy dose of coercion a control to keep workers (and their leaders) from engaging in disloy behavior. In solution 2 labor leaders behave disloyally toward the par Because they are unwilling to violate the loyalty claims of workers, th can be expected to move along the continuum from norm-based voi to exit. Unless the loyalty dilemma is resolved, this solution will bri about the collapse of the alliance.

In solution 3 party leaders behave disloyally toward their own exe utive. Like solution 1, this scenario is consistent with alliance surviv but for very different reasons. Once party leaders defy the executive join labor in resisting the reforms, labor leaders can side with the wor ers without behaving disloyally toward the party and are thus fre from their own loyalty dilemma. Although this solution has the adva tage of building reserves of loyalty, it is unlikely to be sustainable, f two reasons. First, domestic and international pressures are likely limit the party's prospects for convincing the government to aband market reform. Second, most party leaders face formidable organiz tional and electoral obstacles to waging a drawn-out battle against the own executive. As a result, they are likely to try to mend party-exec tive relations, either by cutting deals that involve a compromise wi the reforms or by making internal leadership changes. If this reconci ation occurs without a major change in economic policy, labor leade will once again find themselves in a loyalty dilemma.

I argue that the authority structures in which labor and party leade find themselves shape the choice set among these different solutions. F labor leaders, the most relevant structures are those that determine t relative power of the party and workers to sanction them for disloy behavior. The party's sanctioning power is closely linked to its past a

[26] I recognize that these loyalty dilemmas may not set in immediately when a labor-backed gove ment adopts painful reforms. Like labor leaders, workers are likely to have reserves of loyalty that all the unions to cooperate with the reforms without being viewed as disloyal. Worker patience will w thin, however, if the governing party appears to be betraying its principles and/or the union appears be selling out to the reformers. Moreover, as discussed above, reserves of worker loyalty may alre be quite shallow as a result of the secular decline in class consciousness and collective identities.

present relationship with the state. A hegemonic party that has been in power for many years will control greater resources than a party that has spent decades in exile or must share power with other parties. Likewise, a party with the authority to withdraw benefits already granted will enjoy significantly more sanctioning power than a party restricted to using inducements as a bargaining chip only prior to their delivery. Both of these factors are affected, moreover, by the nature of the state itself. Controlling a weak state in an economy dominated by private interests will bring the party far fewer resources than controlling a strong state capable of imposing its will on domestic and foreign actors.

The sanctioning power of workers is strongly influenced by the dynamics of union competition and, relatedly, by the legal framework governing union formation, collective bargaining, strike activity, union elections, and membership in union organizations.[27] This framework structures the universe of workers ostensibly represented by affiliated labor leaders. Within that universe it affects the extent to which these workers can mobilize independently, remove leaders within the union, switch allegiances to a competing union, and control inducements that labor leaders may want to deliver to the party, such as wage restraint, industrial peace, or working-class votes. To some degree, the sanctioning power of workers varies inversely with that of the party, since a party empowered to intervene extensively in union affairs is likely to constrain the independent actions of workers.

For party leaders the most relevant authority structures are those that determine their capacity to act autonomously vis-à-vis their own executive. If they are in a position to challenge the executive when their party is in government, they may be able to stake out an independent position on behalf of workers and perhaps even block the government's reforms through legislative action. Likewise, if there are opportunities for effective dissent within the party, labor leaders may be able to foment internal opposition to the reforms. Both options are likely to depend on factors such as leadership selection mechanisms, rules governing party factions, sources of party financing, and disciplinary procedures. Although party leaders may not take advantage of authority structures that grant them autonomy from their own executive, such structures must exist for them to join labor in resisting the reforms.[28]

[27] For an excellent discussion of the impact of union competition on labor's responses to market reform, see María Victoria Murillo, "From Populism to Neoliberalism: Labor Unions and Market-Oriented Reforms in Argentina, Mexico, and Venezuela" (Ph.D. diss., Harvard University, 1997).

[28] Factors such as economic conditions, international pressures, ideological proclivities, internal party politics, and electoral dynamics are likely to shape the propensity of autonomous party leaders to challenge their own executive.

AUTHORITY STRUCTURES IN MEXICO, SPAIN, AND VENEZUELA

The three cases presented here suggest that different combinations (authority structures are likely to produce different solutions to the lo] alty dilemmas created by the party's adoption of market reform. I Mexico a party-biased distribution of sanctioning power led to allianc survival despite the party's lack of autonomy from the executive (solu tion 1). In Spain a worker-biased distribution of sanctioning pow combined with the party's lack of autonomy from the executive to prc duce alliance collapse (solution 2). Finally, in Venezuela an intermediat distribution of sanctioning power combined with party differentiatio from the executive to produce alliance survival (solution 3). These dif ferent authority structures affected the relative vulnerability of disaf fected labor leaders to sanctions for disloyal behavior and the party range of options for retaining their loyalty during the reform period.

MEXICO: EXECUTIVE CONTROL AND PRESSURE FROM ABOVE

In Mexico the lines between the party and the executive were s blurred that Mexicans often spoke of "the PRI-government" (*el PRI-go bierno*). Although the PRI was formally empowered to select its ow leaders, set a political agenda, and nominate candidates for public of fice, it "never developed an institutional structure independent of th state apparatus."[29] The chief executive, who was always a PRI candidate served as supreme commander of both the state and the party durin his six years in office (*sexenio*).[30] He chose a significant share of th party's leaders and could remove any PRI bureaucrat or politician wh challenged his authority. More importantly, each outgoing presiden designated his successor, who was then formally nominated by the PR This system gave PRI leaders and their clients tremendous incentives t curry favor with the president, particularly toward the middle of eac *sexenio*.

Executive dominance merged with the centralization of authorit within the PRI. In the 1940s the party adopted reforms that reduced th capacity of organized sectors to nominate their own candidates, height ened the influence of party bureaucrats, and imposed strict party disci pline on PRI deputies and senators.[31] In addition to generating th

[29] Kevin Middlebrook, *The Paradox of Revolution* (Baltimore: Johns Hopkins University Pres 1995), 28.

[30] Mexican presidents (and all other elected officials) are constitutionally prohibited from runnin for reelection.

[31] Pablo González Casanova, *El estado y los partidos políticos en México* (México, D.F.: Ediciones Er 1981), 122–29; Guadalupe Pacheco, "La XIV Asamblea Nacional del PRI," *Estudios Políticos* 3 (1991), 77–78.

jority of legislative proposals, the chief executive dictated the party e in congress. Divisions existed within the PRI, but harsh penalties couraged the public airing of internal disputes. Combined with the 's dominance in the legislative branch, this discipline meant that sidential initiatives almost always became policy. If party actors ed to alter these initiatives through internal negotiations, they had le choice but to throw their support behind the final outcome, howr undesirable.

These blurred lines between party and state reinforced a top-down tribution of sanctioning power that favored the PRI over workers. t only did the PRI dominate all three branches of government for rly seventy years, but Mexico's Federal Labor Law (LFT) gave the te extensive authority to intervene in labor affairs. Labor leaders ld not form a union or hold a strike without the approval of the or authorities, and the PRI-controlled government had the discreiary authority to manipulate union elections and to influence the come of disputes regarding a union's title to a collective contract.

n addition, the CTM's organic yet asymmetrical ties to the PRI enved the party with significant power relative to its labor allies. Been the late 1930s and the early 1950s, the CTM constructed formal l exclusive linkages with the PRI that generated leadership privileges also raised the costs of challenging the party's authority. PRI leaders, anwhile, cultivated multiple and competing allegiances with other cial unions (as well as with peasant and middle-class organizations) t enabled them to employ divide-and-conquer tactics against the M without encouraging dissidence or abandoning their overall comment to labor.

At the same time Mexican workers were seriously constrained in ir ability to punish labor leaders for disloyal behavior. As long as M leaders kept a few lieutenants happy with opportunities for peral gain and patronage, they did not have to worry too much about ng control of their unions. The LFT allowed for—but did not rere—the creation of a closed shop through the incorporation of exsion clauses in collective contracts. An entry clause made union mbership a prerequisite to employment, and a separation clause rered the employer to dismiss any worker who lost his or her union mbership. By linking union membership to employment, these ises effectively protected incumbent labor leaders from dissent from ow, even in the event that workers became disgruntled with their lership or with union policy.

Further, labor leaders were granted wide latitude in managing the in-

ternal affairs of their union. For example, they faced no legal require-ments regarding the use of secret ballots in union elections, the fre-quency of union assemblies, or the structure of union authority.[32] Whil consistent with democratic unionism, this latitude allowed labor leader to impose controls on workers. In fact, the PRI encouraged such prac-tices, preferring to delegate control to affiliated labor leaders rathe than to exercise it directly. On the rare occasions that worker unres overwhelmed these internal authority structures, the PRI used its lega and coercive powers to silence dissenting workers or leaders.

This party-biased distribution of sanctioning power was self-rein-forcing. First, the relative autonomy of CTM leaders from workers en-couraged them to value leadership privileges such as party posts an elected office independently of their impact on workers. This patter was set during an episode of alliance stress in the late 1940s, when th CTM's dominant leadership group rejected a petition by the confedera-tion's former president, Vicente Lombardo Toledano, to support a nev party more sympathetic to working-class interests. Rather than view-ing this alternative as an opportunity to realize the ideological agend of the labor movement, these leaders saw it as a threat to the politica prerogatives that came with their alliance with the PRI. The party re-warded them generously for their loyalty, but at the price of weakenin their legitimacy with workers. In the process, they increased their de-pendence on the PRI by limiting the option of turning to workers fo support in the event that the party abandoned them.[33]

Second, CTM leaders could get away with behaving disloyally towar workers only as long as they remained in the good graces of the PRI Employers tended to negotiate exclusion clauses only with "reliable unions favored by the ruling party. Moreover, the provisions of the LF and the party's dominance of the political system enabled the PRI to ex-pose previously protected labor leaders to voice and exit by workers. In-stead of denying registration to a competing union, PRI authoritie could grant it. Or instead of quelling dissent within the workplace, the could cultivate it. Although the PRI used these tactics almost exclusivel against the CTM's rivals until the 1980s, their successful application sen the message that friends prosper while foes perish. Thus, the threat o worker dissent worked paradoxically to reinforce the CTM's loyalty t the PRI at the expense of its loyalty to workers.

[32] Middlebrook (fn. 29), 66–67.

[33] It should be recognized, however, that the CTM's crisis of legitimacy in the 1940s was ameliorate in the 1950s and 1960s when the PRI's strategy of "stabilizing development" enabled the CTM to delive positive gains in real wages, nonwage benefits, and employment to its members. See Middlebrook (fn 29), 120–22.

Faced with a loyalty dilemma in the 1980s and early 1990s, CTM leaders were encouraged by these arrangements to sacrifice the loyalty claims of workers (solution 1). Although the PRI refused to soften its reforms, which imposed painful sacrifices on the rank and file, CTM leaders did not move beyond norm-based voice. To some degree, they were held hostage by the PRI's divide-and-conquer tactics. Throughout the reform period the Labor Ministry facilitated defections by workers to rival confederations more willing to cooperate with the party's policies.[34] In addition, President Carlos Salinas used charges of corruption to displace intransigent CTM leaders and cultivated relations with rival leaders whom he hailed as representatives of a "new unionism" better suited to Mexico's new economic model.

The unwillingness of CTM leaders to move beyond norm-based voice did not stem entirely from a fear of reprisals, however. Although they were frustrated by the PRI's failure to provide generalized benefits for workers, they continued to count on the party respecting their core interests—which emphasized leadership privileges rather than worker well-being. This preference ordering, which reflected their historical experience of relative autonomy from workers, enabled the PRI to retain their loyalty without seriously compromising its reform program. Rather than having to restore the living standards of workers, the PRI could respond to norm-based voice by providing selective benefits to CTM leaders and their lieutenants.

During the reform period the PRI supported the creation of union-owned businesses, incorporated CTM leaders into high-level wage negotiations, preserved the CTM's quota of party posts and candidacies, delivered on a promise not to reform the LFT, and abandoned plans to eliminate the union's control over the distribution of worker housing. Although some of these concessions trickled down to CTM members, the universe of beneficiaries was narrow, and the gains paled in comparison with the declines in wage and nonwage benefits experienced by most workers during the reform period.[35] Moreover, the last three con-

[34] For a discussion of these tactics, see Javier Aguilar García and Lorenzo Arrieta, "En la fase más aguda de la crisis y en el inicio de la reestructuración o modernización," in Javier Aguilar García, ed., *Historia de la CTM, 1936–1990*, vol. 2 (México, D.F.: Instituto de Investigaciones Sociales, UNAM, 1990); and María Xelhuantzi López, "Reforma del Estado Mexicano y sindicalismo" (M.A. thesis, Universidad Nacional Autónoma de México, 1992).

[35] In addition, several of these benefits proved short-lived or illusory. Many of the CTM-owned businesses established in the mid-1980s were later dismantled as part of the PRI's privatization program. Meanwhile, the economic cabinet set wage levels and and then delivered them to the CTM for ratification. Although real wages made a partial recovery in the early 1990s, the real minimum wage continued its precipitous decline, and real wages failed to keep pace with productivity gains and in 1992 were still at only 61 percent of their 1980 value. Carlos García Villanueva and Alfred Stoll, "Las relaciones

cessions did not provide any new resources to the CTM. Instead, the PRI merely retreated from an all-out assault on the CTM's political and organizational prerogatives.

Had CTM leaders believed that they were destined to lose these prerogatives on top of the hardships imposed on workers, they might have risked resorting to norm-breaking voice. But based on previous episodes of alliance stress, they expected the PRI to close ranks with them in preparation for the presidential succession. As in previous *sexenios*, their loyalty was rewarded when the PRI took steps to repair its relations with the CTM as the succession approached. In 1993 and 1994 the CTM enjoyed a revival of its hegemonic position in the labor movement, a restoration of key privileges within the party, and some valuable concessions in the government's reform of the worker housing institute. In recognition of this historic pattern of reconciliation, the CTM's secretary of social security affirmed that "it is very heart-warming that when the time of the presidential succession approaches, the organic alliance between the [traditional] sectors of the PRI and the government is reestablished."[36]

SPAIN: EXECUTIVE AUTHORITY AND PRESSURE FROM BELOW

Like the PRI, the PSOE lacked autonomy from its own executive. Although it had an institutional structure distinct from the state, party leaders were constrained from acting independently when the PSOE was in power. Between 1982 and 1996 Felipe González served as president of both the party and the government in a parliamentary system with closed electoral lists and strict party discipline.[37] As chief executive, his powers of appointment and dismissal extended to the cabinet, hundreds of joint government and civil service posts, the civil governors in each province, the government delegates to the autonomous communities, and the heads of the Council of State and the National Industrial Institute (INI).[38] He also controlled "the most important decisions re-

laborales en México y su contexto económico y social," in Manfred Wannöffel, ed., *Ruptura en las relaciones laborales* (México, D.F.: Fundación Friedrich Ebert, 1995), 221–24.

[36] Quoted in *El Financiero*, May 7, 1993. Another CTM official drew an explicit parallel between the reaccommodation in the 1990s and the shift by President Luis Echeverría in the 1970s from support for independent unionism to renewal of the alliance with the CTM. Confidential author interview, Mexico City, November 23, 1994.

[37] PSOE statutes required members of the socialist group in parliament to adhere to party discipline regarding their legislative action and votes. PSOE, *Anexo del acta del 29 Congreso* (1981), 52. The party leadership also deducted stiff fines from the paychecks of deputies who missed parliamentary sessions and committee meetings, thereby discouraging abstention as a form of silent protest against the government. Donald Share, *Dilemmas of Social Democracy* (New York: Greenwood Press, 1989), 125.

[38] Peter J. Donaghy and Michael T. Newton, *Spain: A Guide to Political and Economic Institutions* (Cambridge: Cambridge University Press, 1987), 66.

ıting to Government-Parliament relations which may threaten a cabıet's existence—such as the decision to ask for a vote of confidence, to issolve the [parliament], or to call a referendum."[39]

As in Mexico, this executive power merged with a centralization of ıthority within the PSOE. Aware of the debilitating consequences of arty factionalism in the past, a group of socialists led by González ushed through a reform of the party's internal structure at the PSOE's 'wenty-eighth Congress in May 1979. The new rules established an ıdirect, strictly majoritarian, winner-take-all electoral system for deleates to party congresses and shifted voting on congressional resoluons from local groupings to regional blocs.[40] The new rules also :inforced the party's intolerance for dissent, stipulating that a party ıember could be expelled for making public declarations damaging to ıe party image, for serious acts of indiscipline, for slandering another arty member, or for provoking a major conflict within the party.[41] Vhen combined with the power of the chief executive, the PSOE's low :vels of membership, and Gonzalez's popularity among the electorate, ıese rules severely limited the capacity of party actors to block the iniatives of their own government.

In stark contrast to Mexico, however, this lack of party autonomy 'om the executive coexisted with a worker-biased distribution of sanconing power. Besides having been out of power for most of its history, ıe PSOE enjoyed little authority to intervene in union affairs even after 'inning an absolute majority in the parliament in 1982. Breaking with ıe system of industrial relations established under the dictatorship of rancisco Franco, the 1978 Constitution and the 1980 Workers' tatute severely limited state control of union formation, collective barıining, strikes, and leadership selection. The only real leverage PSOE aders had over their labor allies was their control of party posts and ındidate placement on closed lists for elected office.

At the same time, workers had significant power over UGT leaders ırough their participation in factory council elections. The 1980 Vorkers' Statute established two legal mechanisms of worker represen.tion on the shop floor: the union section and the factory council. The ıion section consisted exclusively of union members who acted as "the ıtended arm of unions in the enterprise."[42] By contrast, all workers in

[39] Antonio Bar, "Spain," in Jean Blondel and Ferdinand Müller-Rommel, eds., *Cabinets in Western ırope* (London: Macmillan, 1988), 112.
[40] Share (fn. 37), 56.
[41] Richard Gillespie, *The Spanish Socialist Party* (Oxford: Clarendon Press, 1989), 346–47.
[42] Modesto Escobar, *Works or Union Councils? The Representative System in Medium and Large Sized anish Firms*, Estudio/Working Paper no. 43 (Madrid: Instituto Juan March, 1993), 37.

the workplace, regardless of whether they belonged to a union, could vote in factory council elections. If no union gained an absolute majority of seats, the factory council became the negotiating agent, as occurred in most large firms.

Within the workplace a union's performance in factory council elections determined access to leadership positions and the competencies of the union section. Beyond the workplace, this performance affected (1) the right to negotiate provincial, regional, or national agreements; (2) the legal capacity to convoke council elections in any workplace; (3) representation on the boards of public agencies; and (4) access to government subsidies. Spanish labor law reserved these privileges for unions that qualified as "most representative" by receiving more than 10 percent of all delegates at the national level or more than 15 percent of all delegates at the regional level. Until 1994 the last two privileges were distributed proportionally among these unions on the basis of their electoral performance. Although the UGT was never at any risk of losing its status as a "most representative" union, it faced stiff competition from the CCOO over institutional representation and government subsidies.

The vulnerability of UGT leaders to worker sanctions was heightened by the fact that *all* workers in the workplace could vote in factory council elections.[43] As Ludolfo Paramio argues, "[If] the electoral system is competitive, and the strength and resources of each union depends on its capacity to capture votes, it is easy to understand that union strategies will be strongly influenced by the perceptions of their potential electors."[44] When combined with weak levels of union militancy, this "voters' unionism" reduced the capacity of UGT leaders to retain worker loyalty through the distribution of selective incentives and made them vulnerable to opposition voting.

As in Mexico, the distribution of sanctioning power was self-reinforcing but in the opposite direction. The extensive capacity of Spanish workers to punish UGT leaders for disloyal behavior not only encouraged these leaders to define their core interests in terms of worker well-being but also devalued the few levers of control held by the PSOE. In 1987 several UGT leaders resigned their parliamentary seats rather than

[43] Although only 10–15 percent of salaried workers were union members in the late 1980s, about 50 percent voted in factory council elections, a figure that increased to 80 percent in workplaces where elections were held. Escobar (fn. 42), 6; Faustino Miguélez Lobo, "Las organizaciones sindicales," in Faustino Miguélez and Carlos Prieto, eds., *Las relaciones laborales en España* (Madrid: siglo veintiuno editores, 1991), 220.

[44] Paramio, "Los sindicatos y la política en España, 1982–1992," in Alfonso Guerra and José Felix Tezanos, eds., *La decada del cambio* (Madrid: Editorial Sistema, 1992), 531–32.

vote in favor of a government budget they viewed as unnecessarily draconian. Aware of the potential backlash from below, they perceived these leadership privileges as *liabilities* rather than assets once they conflicted with the defense of worker interests.[45] Their willingness to forgo such privileges, which was inconceivable in Mexico, left the PSOE with few options for retaining the loyalty of union leaders other than a major retreat from its reforms.

In addition, the PSOE's limited capacity to withdraw benefits once granted contributed to a perverse cycle of concessions leading to radicalization. In 1985 and 1986 the socialists bolstered union rights in the workplace, mandated the return of property confiscated from the unions by the Franco regime, and guaranteed union representation on the governing boards of public agencies. Besides meeting several of the UGT's long-standing demands, these measures privileged the UGT over its rivals. But because the PSOE could not easily withdraw these concessions once granted, their provision ended up enhancing the UGT's freedom to make more radical demands. As a former PSOE labor minister complained, "We gave them the guns and the ammunition to wage their struggle effectively and they turned around and used them against us."[46]

These authority structures encouraged UGT leaders to sacrifice the loyalty claims of the party when confronted by a loyalty dilemma in the 1980s (solution 2). As evidence mounted that the PSOE would not fufill its historic promise to construct a Keynesian welfare state, their loyal attitude soured into a deep sense of betrayal. Combined with the party's lack of sanctioning power and an apparent protest vote by workers against the UGT in the 1986 factory council elections, this disillusionment accelerated their movement along the continuum from norm-based voice to exit.[47] By the end of 1988 the UGT's distrust of PSOE

[45] UGT leaders were particularly sensitive about being too closely associated with the PSOE's reforms after performing poorly in key firms and sectors in the 1986 factory council elections. Following the elections, Redondo complained that "each time a minister speaks . . . we lose votes; each time the 27 liberalizing measures are mentioned, we lose votes. . . . There is a problem of identification. Workers identify us with the government or with the actions of public managers." Quoted in *El País*, December 18, 1986.

[46] Author interview with Luis Martínez Toval, Madrid, June 29, 1995.

[47] Although it is possible that UGT leaders moved more quickly toward exit than the rank and file would have demanded, the timing of their movement suggests they were responding, at least in part, to signals from workers. Moreover, several opinion polls suggest that workers sympathized with the UGT's criticism of the government. See, e.g., Fermín Bouza et al., *Perfil, actitudes y demandas del delegado y afiliado a UGT* (Madrid: Fundación Largo Caballero, 1989), 252–53; and *REIS* 38 (April–June 1987), 358–59. For a good discussion of the radicalism of Spanish labor leaders relative to the rank and file, see Robert Fishman, *Working-Class Organization and the Return to Democracy in Spain* (Ithaca, N.Y.: Cornell University Press, 1990), 39–45.

leaders was so great that only a radical policy change by the governmer could have salvaged the alliance.

This situation might have been avoided had the party been able an willing to take the UGT's side against the reforms. But the opposite hap pened. Although the UGT won the party's support in several of its bat tles with the government between 1982 and 1984, this was onl because González agreed to compromise. When both sides dug in the heels after 1985, the party closed ranks with the government. The la straw for UGT leaders was a campaign led by the party's organization mastermind, Alfonso Guerra, to foment divisiveness within the UG against the 1988 general strike.[48] Guerra's transformation from ally t opponent heightened the UGT's sense of betrayal, reinforced its grow ing perception of the party as a coconspirator with the government, an eliminated any remaining possibility that sympathizers in the part might be able to soften the government's policies.[49]

VENEZUELA: PARTY AUTONOMY AND PRESSURE FROM BOTH SIDES

In contrast to both Mexico and Spain, party-executive authority struc tures in Venezuela gave AD leaders enough autonomy to challenge thei own government. Supreme authority within the party resided with th National Executive Committee (CEN) rather than with the chief exec utive. Once in office, an AD president had no formal position in th party executive and exercised no direct control over the party appara tus. The CEN, by contrast, controlled the placement of party candidate on closed lists for the Senate and the Chamber of Deputies, preselecte the head of the party's parliamentary fraction, dictated the party line i congress, and even had final say over when an AD president could forr a multiparty coalition to pass legislation. Moreover, the president ha only limited influence over the nomination of his successor.[50] In con trast to Mexico, the party's nominating convention enjoyed both for mal and de facto power to select its presidential candidate, and it delegates often voted down the president's choice.

These spaces for autonomous action enabled AD leaders to join th CTV in resisting the government's reforms at a critical moment i

[48] A year earlier, the PSOE had refused to defend a budget compromise negotiated by Guerra bu then rejected by González. This outcome prompted one UGT leader to complain that "the party ha been suppressed in the government-party relationship." Quoted in *Cambio 16* (October 26, 1987).

[49] Although members of a critical tendency in the PSOE, the socialist left, sympathized with th UGT's position, they lacked any real power within the party because of electoral rules that placed cor trol primarily in the hands of the central party apparatus.

[50] As in Mexico, the Venezuelan president is constitutionally prohibited from succeeding himself, a though he can run for reelection after being out of office for at least ten years.

party-union relations (solution 3). Not long after CTV leaders organized the 1989 general strike, AD eased their loyalty dilemma by defying Pérez and behaving much like an opposition party.[51] AD's bitter differences with the executive broke into the open after the party's disappointing showing in state and local elections in December 1989. The AD leadership publicly blamed the results on the economic reforms and the lack of AD representation in the cabinet. The struggle became increasingly vitriolic in 1990 and 1991, culminating in a landslide victory by an anti-Pérez faction, led by Luis Alfaro Ucero, at AD's national convention in October 1991.

This feud had two salutary effects on CTV leaders. First, it effectively freed them to treat the Pérez administration like a non-AD government, thereby transforming actions that would normally constitute norm-breaking voice into norm-based voice. As in the years when AD's main rival, COPEI, was in power, the CTV got the green light to support antigovernment initiatives in congress, carry out joint actions with other labor confederations, and mobilize workers. Second, AD's congressional delegation took the CTV's side against the executive on several important issues. In 1990 AD supported a labor law enhancing worker rights and union privileges despite strong opposition by Pérez and his cabinet. In 1991 AD voted down an executive decree establishing a salary increase below the rate of inflation and suspended deliberations regarding a controversial bill offered by the executive to rationalize severance benefits.

These developments brought CTV leaders important relief from the loyalty dilemma created by Pérez's economic reforms. Besides being able to resist the reforms without engaging in disloyal behavior toward the party, they received crucial help from the party in postponing reforms that would have put even greater strain on their relations with workers. This relief arrived, moreover, at a critical moment in party-union relations. By the time outright confrontation between AD and the Pérez administration subsided in 1992, the momentum had shifted back to norm-based voice. A pro-Pérez leader had become president of the CTV, and two attempted military coups had diverted attention from the government's reforms. Furthermore, the episode reaffirmed the possibility that CTV leaders could find allies within the party in their struggle against the reforms.

Had CTV leaders been highly vulnerable to pressures from below, temporary access to solution 3 might not have been sufficient to salvage

[51] Javier Corrales, "El presidente y su gente," *Nueva Sociedad* 152 (1997), 97.

the alliance. But instead they faced an intermediate distribution of sanctioning power. On the party side AD had a significant degree of leverage but not as much as the PRI. After a decade in exile AD controlled the presidency for twenty-five of the thirty-five years following the democratic transition in 1958 and always held the largest share of seats in the congress. In addition the 1936 Labor Law, which was modeled after Mexico's LFT, gave the state extensive authority to intervene in union formation, collective bargaining, strikes, and union elections. Furthermore, AD's organic ties with individual unionists translated into significant influence over leadership selection within the labor movement. The party decided which AD labor leaders would be placed on candidate lists for union elections; these positions belonged to party in the sense that it could replace a suspended or expelled leader.[52] Thus AD unionists depended on the party not only for party posts or elected office but also for their upward mobility within the union.

Despite these powerful levers of control, AD's sanctioning power was constrained by two factors. First, unlike the PRI (or the PSOE in the 1980s), AD had to negotiate the terms of its hegemony with other parties. Both the political system and the labor movement operated on the basis of power sharing and proportional representation among the major parties, which meant that AD could not act with impunity in either arena. Second, labor leaders enjoyed a degree of influence within the party that their counterparts in both the CTM and the UGT lacked. Like the CTM (and in contrast to the UGT), AD's Labor Bureau constituted the most unified and organizationally autonomous grouping within the party.[53] But unlike the CTM, the Labor Bureau could translate this unity into pivotal influence over leadership selection within the party.[54] As a result, party leaders had strong incentives to win the loyalty of the Labor Bureau, particularly during periods leading up to internal elections.

On the worker side, CTV leaders enjoyed significant protections from worker dissent but were more vulnerable than their counterparts in Mexico. A combination of indirect elections, negotiated slates, disproportionate voting power for small unions, and "delegates for life" protected high-level leaders from the kind of voters' unionism that

[52] León Arismendi and Francisco Iturraspe, "Sistemas electorales sindicales," in Manuel Vicente Magallanes, ed., *Mandato Político, Evolución Electoral, Comunicación y Sociedad* (Caracas: Publicaciones del Consejo Supremo Electoral, 1990), 257.

[53] UGT members constituted a sizable share of the delegates to PSOE congresses, but the UGT did not act as an organized faction within the party.

[54] Michael Coppedge, *Strong Parties and Lame Ducks* (Stanford, Calif.: Stanford University Press, 1994), 109.

prevailed in Spain. At the level of the firm, however, Venezuelan workers had real if limited access to both voice and exit. The Venezuelan congress removed exclusion clauses from an early draft of the 1936 Labor Law, and parallel unions were widespread, particularly in the public sector. In addition, nonunionized workers had the right to bargain collectively if they represented at least 75 percent of the workers in the firm. These features gave workers the option of exiting from a union that did not serve their interests.

More importantly, party pluralism made if possible for workers to punish incumbent leaders by supporting challengers from another party. Although the parties continued to negotiate united slates at the federal and confederal levels into the 1980s, this practice became less and less common at the level of the firm. As a result AD faced serious challenges to its dominance in strategic sectors of the economy. In 1987 it lost control of Venezuela's largest union (SUTISS, located in a state-owned steel complex in the industrial state of Bolívar) to the Radical Cause Party.[55] Although a one union–one vote rule prevented SUTISS and other radical unions from gaining seats on the governing boards of higher-level organizations, AD faced the possibility of losing its majority on the CTV executive if enough firms came under the control of rival parties. This threat put additional pressure on AD labor leaders to outbid these parties for worker support.

This intermediate distribution of sanctioning power meant that CTV leaders were likely to suffer serious consequences regardless of which claimants they betrayed in a loyalty dilemma. As a result, they tended to vacillate between different strategies and to consider both leadership privileges and generalized benefits for workers as part of their core interests. But as long as they could manage occasional demonstrations of loyalty toward workers, their preference was to remain in the good graces of the party. From 1984 to 1988 AD accommodated this preference by combining high-level positions for CTV leaders with benefit packages for workers and expansionary spending just before the 1988 elections. These concessions kept the CTV from moving beyond norm-based voice despite the government's brief experiment with austerity and a secular decline in the living standards of workers.

The CTV's balancing act demanded more radical action, however, after Pérez's reforms prompted violent demonstrations in the streets of Venezuela's major cities in February 1989. This unanticipated burst of

[55] For a detailed discussion of the struggle for control of SUTISS, see Steve Ellner, *Organized Labor in Venezuela, 1958–1991* (Wilmington, Del.: Scholarly Resources, 1993), 153–62.

opposition by the popular sectors dealt a serious blow to the CTV's legitimacy and emboldened those leaders within the CTV who favored a more combative stance toward the government. The CTV president, Juan José Delpino, concluded that "if we do not maintain conduct that is autonomous, independent, critical, rebellious toward the parties and the government, then this labor movement will escape our control."[56] Against the explicit wishes of the party and the government, the CTV engaged in an unprecedented act of norm-breaking voice by joining with rival labor confederations to organize the May 1989 general strike.

Having restored some of their lost credibility, CTV leaders quickly retreated to norm-based voice despite Pérez's refusal to abandon his reforms.[57] In part, they were resorting to their established practice of trying to appease competing claimants by vacillating between alternative strategies. But their retreat is unlikely to have been so rapid or complete had the party not offered them relief from the loyalty dilemma in 1990 and 1991. Rather than having to defy the party to demonstrate their loyalty to workers, as in May 1989, they could now do so with the party's blessing and support. Although this window of opportunity eventually closed, it proved sufficient to swing the pendulum back toward norm-based voice for the duration of the Pérez administration.

CONCLUSION

The revised version of exit, voice, and loyalty offered in this article provides a theoretical framework for understanding the fate of party-union alliances that come under stress when the party adopts market reform. At the heart of this framework is the concept of loyalty. Understood as historically constructed expectations and adherence to codes of conduct, loyalty is built into the principles and norms on which these alliances are based. When a labor-backed government adopts reforms that harm workers and unions, it challenges these principles and norms and creates a loyalty dilemma for both labor and party leaders.

Initially, the reserves of loyalty accumulated prior to the party's adoption of reforms are likely to inhibit disaffected labor leaders from moving beyond norm-based voice in response to a loyalty dilemma. If the government persists with the reforms, however, the propensity of these leaders to behave disloyally toward the party will vary according to their relative vulnerability to reprisals for disloyal behavior, and the

[56] Quoted in *El Nacional*, April 6, 1989.

[57] Delpino resigned in frustration from his post as CTV president, and the CTV began cooperating with Perez's privatization policies in return for selective benefits for laid-off workers.

party's capacity to retain their loyalty even as sacrifices are imposed o workers. Both of these variables are linked to the authority structure in which labor and party leaders find themselves. Assuming the gov ernment does not significantly modify its reforms, the alliances mos likely to survive are those in which labor leaders face a party-biased dis tribution of sanctioning power and/or the party is able and willing t differentiate itself from its own executive to join labor in resisting th reforms.

Alliance survival during the transition to a market-led economy ma not be the end of the story, however. Both Mexico and Venezuela ar in the midst of political crises that threaten the viability of their party union alliances. These crises are linked, morever, to the same variabl that allowed for alliance survival during the 1980s and 1990s: the au tonomy of political elites from their bases. In both countries popula frustration with the inadequacy of interest representation is fueling th collapse of the traditional party systems. An important consequence c this collapse is a redistribution of sanctioning power away from the pai ties. In Mexico the PRI faces growing electoral competition and can n longer count on controlling all three branches of government. I Venezuela AD is in danger of becoming completely marginalized by th victory of independent candidates in presidential elections and the con vening of a constituent assembly to rewrite the Venezuelan constitutior

Initially, affiliated unions in both countries responded to these crise by closing ranks with their party allies, since the potential cost of dis loyalty now included the collapse of the power structures on which the had come to depend. Ultimately, however, they are likely either to adap by turning to more flexible and less partisan strategies or to be over taken by rival organizations that have already done so. Indeed, thes patterns are already emerging. In Mexico CTM leaders continue to clin to the PRI, but they are losing ground to new labor organizations com mitted to greater autonomy from political parties and the state. I Venezuela CTV leaders have been more proactive about delinking fror AD, but it may too little too late to save the confederation from bein completely reorganized by antiparty reformers.[58]

These developments suggest that party-union relations in Mexicc Spain, and Venezuela may be taking alternative paths to a similar des tination of more contingent, ad hoc, localized, and loosely institution alized patterns of interaction. This outcome would be consistent wit

[58] For a more detailed analysis of these trends, see Katrina Burgess, "Interest Representation at Critical Juncture? Unions and Parties in a Liberalizing World" (Prepared for delivery at the annu meeting of the American Political Science Association, Atlanta, Ga., September 2–5, 1999).

broader trends in interest representation. Greater fluidity in the economic sphere seems to be encouraging greater fluidity in the political sphere. Thus, the old pattern of stable, institutionalized linkages informed by strong bonds of loyalty may be giving way to new patterns characterized by shifting allegiances, shorter time horizons, fragmented forms of collective action, and weaker loyalties.[59] As we have seen, the degree of political upheaval required to make this transition varies significantly across countries. In Spain labor leaders delinked from their party allies incrementally and within existing structures of authority. In Mexico and Venezuela, by contrast, it is taking crises of profound proportions to unravel old linkages and loyalties—crises whose final impact on authority relations has yet to be determined.

[59] For informative discussions of the changing nature of interest representation, see Russel J. Dalton, Paul Allen Beck, and Scott C. Flanagan, "Electoral Change in Advanced Industrial Democracies," in Dalton, Beck, and Flanagan, eds., *Electoral Change in Advanced Industrial Economies: Realignment or Dealignment?* (Princeton: Princeton University Press, 1984); Douglas A. Chalmers, Scott B. Martin, and Kerianne Peister, "Associative Networks: New Structures of Representation for the Popular Sectors?" in Douglas A. Chalmers et al., eds., *The New Politics of Inequality in Latin America* (Oxford: Oxford University Press, 1997); and Frances Hagopian, "Democracy and Political Representation in Latin America in the 1990s: Pause, Reorganization, or Decline?" in Felipe Agüero and Jeffrey Stark, eds., *Fault Lines of Democracy in Post-Transition Latin America* (Miami, Fla.: North-South Center Press, 1998).

ABSTRACTS

FROM POPULISM TO NEOLIBERALISM

LABOR UNIONS AND MARKET REFORMS IN LATIN AMERICA

By M. VICTORIA MURILLO

In the late 1980s, populist labor parties, which had advanced protectionism and state intervention in the postwar period, implemented market-oriented reforms in Argentina, Mexico, and Venezuela. In all three countries, market reforms hurt their union allies. The interaction between allied unions and governing labor parties, however, varied across countries and across sectors within the same country. While some unions endorsed neoliberal reforms, others rejected them despite their long-term alliance with governing parties. While some unions obtained concessions, others failed to do so.

This article argues that the incentives created by partisan loyalties, partisan competition, and union competition explain these interactions. Partisan loyalty results from the long-term affiliation of unions with a political party. Partisan competition takes place among union leaders affiliated with different political parties for the control of unions. Union competition occurs in diverse national and sectoral contexts among labor organizations for the representation of the same workers. Loyalty derived from a long-term affiliation with the incumbent party facilitates collaboration between labor unions and the government. Yet, if partisan competition makes loyal nion leaders afraid of being replaced by activists affiliated with the opposition parties, their inentives for militancy increase as a way of showing their responsiveness to the rank and file hurt y market reforms. Union competition for the representation of the same workers makes coorlination more difficult, thereby weakening unions and making them less likely to obtain conessions from the government despite their partisan loyalty. The article presents empirical vidence from eighteen cases, including national confederations and individual unions in five conomic sectors in Argentina, Mexico, and Venezuela to test this theory.

CONTINGENT DEMOCRATS

INDUSTRIALISTS, LABOR, AND DEMOCRATIZATION IN LATE-DEVELOPING COUNTRIES

By EVA BELLIN

Many classic works of political economy have identified capital and labor as the champions of lemocratization during the first wave of transition. By contrast, this article argues for the coningent nature of capital and labor's support for democracy, especially in the context of late deelopment. The article offers a theory of democratic contingency, proposing that a few variables, amely, state dependence, aristocratic privilege, and social fear account for much of the variation ound in class support for democratization both across and within cases. Conditions associated vith late development make capital and labor especially prone to diffidence about democratizaion. But such diffidence is subject to change, especially under the impact of international ecoomic integration, poverty-reducing social welfare policies, and economic growth that is widely hared. Case material from Korea, Indonesia, Mexico, Zambia, Brazil, Tunisia and other counries is offered as evidence.

THE RELIGIOUS ROOTS OF MODERN INTERNATIONAL RELATIONS

By DANIEL PHILPOTT

The Protestant Reformation was a crucial spring of modern international relations. Had it ever occurred, a system of sovereign states would not have arrived, at least not in the form or at he time that it did at the Peace of Westphalia. This is the counterfactual the author seeks to susain. He first advances an elaborated but qualified defense of the conventional wisdom that Westphalia is the origin of modern international relations. He then accounts for how Protestant deas exerted influence through transforming identities and exercising social power. Structural heories, emphasizing changes in material power, are skeptical of this account. The author roots is empirical defense of ideas in the strong correlation between Reformation crises and polities'

interests in Westphalia. A description of the historical causal pathways running from ideas to political interest then follows. Germany and France are brought as cases to illustrate two of these pathways. Finally, the author shows the evidentiary weakness of alternative structural material explanations.

THE MANY VOICES OF POLITICAL CULTURE

ASSESSING DIFFERENT APPROACHES

By RICHARD W. WILSON

Works in political culture have employed a variety of approaches with different assumptions and methodologies. Commentary on the field, however, generally fails to take account of these differences or to assess the merits of the different approaches. Recent works that typify five different approaches are evaluated in terms of the ways that they conceptualize preference formation and the linkage between preferences and cultural norms. The tendency in most approaches to conflate preferences and norms points to the need for further theoretical development.

FROM POPULISM TO NEOLIBERALISM

Labor Unions and Market Reforms in Latin America

By M. VICTORIA MURILLO*

A TALE OF THREE CARLOSES

CARLOS Salinas, Carlos Andrés Pérez, and Carlos Menem were inaugurated between December 1988 and July 1989 as presidents of Mexico, Venezuela, and Argentina, respectively. They not only shared the timing of their administrations[1] but also the task of turning populist labor parties toward neoliberalism. All three had been the candidates of populist labor-based parties that had advanced protectionism and state intervention in the postwar period. Once in office and facing tremendous fiscal crises and capital flight, all three presidents reduced state intervention and opened the economies of their countries. This shift in development strategy not only was the most important policy turnaround of the postwar era in all three Latin American countries, but it also moved their labor-based parties away from the policies upon

*Earlier versions of this paper were presented at the twenty-first meeting of the Latin American Studies Association, Chicago, September 24–26, 1998, and at the conference "Space, Place, and Nation: Reconstructing Neoliberalism in the Americas," University of Massachusetts, Amherst, November 20–21, 1998. I would like to thank Arun Agrawal, Ernesto Cabrera, Ernesto Calvo, Javier Corrales, Miguel Glatzer, Miriam Golden, Pauline Jones-Luong, Robert Kaufman, Steven Levitsky, James McGuire, Nicoli Nattrass, Phillip Oxhorn, Héctor Schamis, Jeremy Sikkins, and four anonymous reviewers for helpful comments to previous versions of this paper, and Antonieta Mercado for her research assistance in Mexico. I would also like to thank Alberto Alesina, Robert Bates, David Cameron, John Coatsworth, Jorge Domínguez, Geoffrey Garrett, Peter Hall, Ian Shapiro, Juan Carlos Torre, and Deborah Yashar, who provided useful comments to clarify the ideas presented in this paper. I gratefully acknowledge the financial support of the Mellon Foundation, the Harvard Academy for International and Area Studies, and the David Rockefeller Center for Latin American Studies at Harvard University, as well as the Institute for the Study of World Politics. I also want to acknowledge the institutional support of the Department of Political Science at Yale University; the Instituto Torcuato Di Tella in Argentina; the IESA in Venezuela; the Colegio de México, Flacso, and ITAM, and Intelmex in Mexico.

[1] Salinas was president of Mexico from the end of 1988 to the end of 1994, and Carlos Andrés Pérez was president of Venezuela from the beginning of 1989 to mid-1993. To hold the international conditions constant in the comparison with Mexico and Venezuela, I analyze the first administration of Carlos Menem in Argentina, which ran from mid-1989 to mid-1995.

which their historic relationships with long-term union allies had been built.

Despite the parallel convergence of labor-based populism turning into neoliberalism and the common challenge faced by unions in Argentina, Mexico, and Venezuela, union responses to neoliberal reforms were diverse. Union-government interactions varied across these countries and across sectors within the same country. For instance, the Mexican Workers' Confederation subordinated to Salinas's policies, endorsing them in corporatist pacts and even campaigning actively for the North American Free Trade Agreement (NAFTA). To the contrary, the Venezuelan Workers' Confederation opposed Perez's reforms by organizing the first economic general strikes in the history of Venezuela. Argentine teachers unsuccessfully resisted the decentralization of schools; their militancy accounted for more than a third of the total strikes in the two years before the government, with no union input, finally implemented the reform. In Mexico, however, teachers successfully opposed key pieces of the decentralization process and limited its scope. Meanwhile, Mexican telephone workers negotiated with their government to support the privatization of their company in return for job stability, handsome social benefits, and easy financing for the purchase of almost 5 percent of the shares in the company.

Such variation in union-government interactions in response to market-oriented reforms has considerable political importance. First, unions can organize support for or opposition to the reforms, thus changing the costs of reform for policymakers. Their actions can affect the feasibility, design, and implementation of reforms, as shown by the delays in the reform of labor market regulations in all three countries. Unions should therefore be included in the analysis of the politics of economic reforms.

Second, unions can have a direct impact on the governance of administrations led by labor-based political parties. Unions not only have organized the core constituency of labor-based parties and provided them with political machines for electoral campaigns, but they also have shared a long-term partisan identity with governing politicians. As a result, the interaction between unions and labor-based parties implementing market-oriented reforms can influence the future of the electoral coalition that brought these parties to power. Argentina, Mexico, and Venezuela are key cases in that regard because the alliance between unions and labor-based parties had shaped the national party system fifty years earlier.[2]

[2] Ruth Berins Collier and David Collier, *Shaping the Political Arena* (Princeton: Princeton University Press, 1991).

Thus, a theory that accounts for the cross-sectoral and cross-national variations in union-government interactions in these countries can help solve the puzzle of the transition from closed to open economies in Latin America.[3] Moreover, an explanation of such variation has broad comparative implications for our understanding of similar stories in other regions of the world where the pressure of globalization turned labor-parties toward policies that challenged their long-standing agreements with labor allies.

Theories of union politics suggest that union behavior varies according to either differences in national-level institutions or interests of economic sectors with regard to market policies and international integration. Yet, in Argentina, Mexico, and Venezuela, the interactions between unions and labor-based governments over market-oriented reforms cut across national borders and sectoral cleavages. Explanations based on either economic interest or national institutions are thus insufficient to understand such patterns of interaction. I argue instead that the incentives created by partisan loyalties, partisan competition, and union competition explain these interactions. Partisan loyalty results from the long-term affiliation of a union with a political party. Partisan competition is the struggle for control of the unions among union leaders affiliated with different political parties. Union competition is the rivalry among labor organizations for the representation of the same workers, which can take place in diverse national and sectoral contexts. Loyalty derived from a long-term affiliation with the incumbent party facilitates collaboration between labor unions and the government. Yet, if loyal union leaders are afraid of being replaced by activists affiliated with the opposition parties because of partisan competition, their incentives for militancy increase to show their responsiveness to the rank and file hurt by market reforms. Union competition for the representation of the same workers makes coordination more difficult thereby weakening unions and making them less likely to obtain concessions from the government despite their partisan loyalty. I use the empirical evidence from eighteen cases, including national confederations and individual unions in five economic sectors in Argentina, Mexico, and Venezuela to test this theory.

The article is divided into six sections. The first section introduces the policy shift of labor-based parties from populism to neoliberalism

[3] The politics of Argentina, Mexico, and Venezuela have a strong effect in Latin America and the Caribbean. By 1995 they made up 32 percent of the regional population, 43 percent of the regional gross domestic product, and 48 percent of regional exports. Inter-American Development Bank, *Economic and Social Progress in Latin America* (Baltimore: Johns Hopkins University Press, 1996), 357–61.

in Argentina, Mexico, and Venezuela. The second examines the relevant literature upon which I build my analysis. The third describes the patterns of union-government interaction in the context of market-oriented reforms. The fourth presents my hypothesis for explaining those patterns. The fifth section reports the test of my hypothesis on the case studies, and the last offers some conclusions.

From Populism to Neoliberalism

On October 17, 1945, a multitude of workers organized by their unions marched into downtown Buenos Aires to demand the release of imprisoned Colonel Juan Perón and to defend the social legislation he had implemented as secretary of labor. In the elections of February 1946, the Labor Party—organized by the leaders of those same unions—led the political coalition that elected Perón to the presidency. Most workers voted for Perón and taught their children to be Peronist, creating one of the strongest partisan loyalties in Latin America. After all, Perón and the Peronist unions had changed their lives by providing better wages and labor benefits, social security, and even paid vacations at union resorts. Since labor benefits compensated them for previous frustration in dealing with indifferent employers and hostile governments, the unions identified with Peronism and served as Peronist political machines. Peronism, which became the symbol of workers' rights, promoted import substitution industrialization and state intervention as development strategies.[4] It also turned unions into key players in the Argentine political system.

In 1989 a Peronist candidate, Carlos Menem, won the presidential elections. During his populist electoral campaign, Menem promised wage hikes and social justice, and he threatened not to pay the external debt. After his inauguration, however, he delivered austerity, followed by trade liberalization, privatization, and adjustment of the public sector. As state intervention had done forty years earlier, the retreat of the state and market-oriented reforms reshaped state-society relations. The national labor confederation, the Peronist-dominated General Confederation of Labor (CGT), although surprised by the policy turnaround, accepted market-oriented reforms and reduced the number of general

[4] Import substitution industrialization (ISI) was a development strategy adopted by most Latin American countries after the Great Depression. They originally raised tariffs to compensate for the shortage of foreign exchange produced by the crisis, but this policy gradually evolved into active protectionism that included subsidized exchange rates for importing inputs with closed markets. ISI created few incentives for developing internationally competitive industries and, thus, exporting.

strikes from thirteen against the previous administration of the Civic Radical Union (UCR) (1983–89) to only one under the first Menem administration (1989–95). Peronist unions, though, were able to negotiate concessions on the reforms of social security, labor legislation, and privatization.

In Mexico during the same period, the story of the policy turnaround of labor parties in government was similar to that in Argentina. Unions entered the political arena during the Mexican Revolution. They first formed six "Red Battalions" that fought with the army of President Venustiano Carranza in 1915. In 1919 they organized a Labor Party, which supported the elections of Presidents Alvaro Obregón and Plutarco E. Calles. In return, unions obtained political appointments and favorable labor legislation to compensate for their weakness in collective bargaining with private employers. In 1936 President Lázaro Cárdenas promoted the organization of the Mexican Workers Confederation (CTM). In 1938 he founded the Party of the Mexican Revolution (PRM) and integrated labor into its functional structure. Unions became political machines for the PRM, which would become the Institutional Revolutionary Party (PRI), and were included in the party structure. Unions also gave Cárdenas a suitable excuse—in the form of a labor conflict—for the nationalization of oil in 1938. In return, workers received social benefits, and union leaders gained political influence they could use in the industrial arena. The PRI also implemented policies of import substitution industrialization while using the state as the main instrument for economic development.

Half a century later, however, PRI labor leaders witnessed a PRI president, Carlos Salinas, become the champion of economic liberalization, state shrinkage, and market-oriented reforms. The main national confederation, the PRI-dominated CTM, explicitly supported Salinas's stabilization plan and structural reforms both in the annual corporatist pacts signed with government and business representatives and in an agreement to increase labor productivity signed in 1992. Moreover, it campaigned for Salinas's proposal to integrate Mexico into NAFTA. In spite of the CTM's acquiescence, the administration ignored most of its demands even while paying lip service to the alliance between the PRI and the labor movement.

During the same period, a similar story of policy turnaround was being written in Venezuela, where Democratic Action (AD) had traditionally been a champion of democracy and nationalism as well as of union organization. During its first administration (1946–48), AD promoted an upsurge in union organization and social mobilization.

Democratic Action union leaders promoted the founding of the Venezuelan Workers Confederation (CTV), and its labor bureau integrated union leaders into the party structure. Unions provided AD with political machines while channeling workers' loyalty and supporting AD development policies based on state intervention and import substitution industrialization. In return, AD administrations provided workers with social and labor benefits, and union leaders with political influence.

In 1988 AD union leaders endorsed the populist Carlos Andrés Pérez in the primaries of AD and in his campaign for the presidency. Pérez extended state intervention, established a minimum wage, and made dismissals more difficult during his first administration in the 1970s. After his inauguration in February 1989, however, he surprised foes and followers with his announcement of the "Great Turnaround." The Great Turnaround included trade liberalization, macroeconomic adjustment, and structural reforms of the state. The response of Perez's union allies to his reforms differed from that encountered by his Argentinean and Mexican counterparts. The AD-controlled CTV responded to his policy shift by organizing the first economic general strike in Venezuelan history, followed by a series of demonstrations and other strikes that boycotted most of Pérez's reforms in the labor and social sectors. Opposition by the CTV resulted in diverse concessions related to labor-market and social-security reforms until the government reformist intentions receded under the pressure of social protests and two failed military coups.

The Policy Turnaround and the Challenge for Labor Unions

These convergent policy turnarounds in Argentina, Mexico, and Venezuela resulted from the failure of strategies based on state intervention and import substitution industrialization. The 1982 debt crisis highlighted their strategic limitations while it worsened fiscal deficits and balance-of-payment difficulties. Macroeconomic instability and recession followed during the 1980s, the "lost" decade. By the end of the eighties, incumbent populist labor-based parties in Argentina, Mexico, and Venezuela had shifted toward market-oriented reforms.[5] Accord-

[5] Market-oriented reform included short-term stabilization measures, fiscal restraint, tax reform, financial liberalization, competitive exchange and interest rates, trade liberalization, privatization, and deregulation of most markets, including the labor market. See John Williamson, *Latin American Adjustment: How Much Has Happened?* (Washington, D.C.: Institute for International Economics, 1990).

ing to Cukierman and Tommasi, labor-based parties have a comparative advantage in implementing market-oriented policies that bring uncertainty to their constituencies because they enjoy their trust.[6] That is, they are more credible when they claim that it was the exogenous shock rather than ideology that had induced them to implement market-oriented reforms. Yet, to bring capital back into their economies, labor-based parties in government have to overcome business's distrust of their previous populist character. Their policies, therefore, should be more drastic to show their new commitment to the market.

Drastic market reforms, though, had costs for labor unions and workers whose influence had developed based on state expansion, protectionism, rigid labor markets, and political clout. Trade liberalization increased differences among workers across and within sectors, making it harder to organize labor unions based on horizontal solidarity. International competition and privatization also provoked labor restructuring and layoffs in sectors that had been among the most highly unionized in the past, thus reducing the relative influence of unions. Higher unemployment hurt union bargaining power and increased job instability for union constituencies.[7] Stabilization policies that relied on wage restraint and international competition further reduced their wage bargaining power.[8] Moreover, the reform of social and labor regulations challenged institutions that had provided unions with legal and political clout—from appointments on social security boards to monopolies of representation—which they would not have been able to achieve based solely on their industrial power. More importantly, mar-

[6] Alex Cukierman and Mariano Tommasi, "Credibility of Policymakers and of Economic Reform," in Federico Sturzenegger and Tommasi, eds., *The Political Economy of Reform* (Cambridge: MIT Press, 1998).

[7] In Argentina, according to the permanent household surveys of the National Institute of Statistics and Censuses (INDEC), unemployment increased from 6.5 percent in 1988 to 18.6 percent in 1995. In Venezuela unemployment rose from 6.9 percent in 1988 to 9.6 percent in 1989 and 10.4 percent in 1990, although it fell to 6.5 percent in 1993. See Keila Betancourt, Samuel Freije, and Gustavo Márquez, *Mercado labora: Instituciones y regulaciones* (Labor markets: Institutions and regulations) (Caracas: IESA, 1995), 5. In Mexico, according to the International Labour Organisation, (ILO), unemployment measurements are highly contentious, but open unemployment peaked in 1983 and 1984. See ILO, *Yearbook of Labour Statistics* (Geneva; ILO, various years). In addition, the combined official rate of open unemployment and underemployment grew from 6.8 percent in 1989 to 8 percent in 1994. See Carlos Salinas de Gortari, *VI Informe de Gobierno* (VI State of the union) (Mexico City: Presidencia de la Nación, 1994).

[8] In Argentina, hyperinflation cut manufacture real wages by 36.3 percent between January 1989 and March 1991. Even after the success of stabilization, manufacture real wages fell by 12 percent between April 1989 and June 1995. Consejo Ténico de Inversiones, *La economía argentina:. Anuario 1997* (The Argentine economy: Yearbook 1997) (Buenos Aires: CTI, 1997), 65. In Venezuela the real industrial wage fell 35 percent in the 1989–93 period. Unido Industrial Statistics Database, 3-digit (1998). In Mexico real wages in manufacturing had dropped by almost 40 percent between 1982 and 1988, and despite improvements during the Salinas administration, they did not recover their 1982 level. ILO (fn 7).

ket reforms introduced a high degree of uncertainty about the future positions of union constituencies in the labor market, which often induced them to reject the changes. All these effects were more acute in the sectors that had previously enjoyed high levels of protection from competition (such as the public and the manufacture sectors).

In spite of the effect of market-oriented reforms on unionized labor, incumbent labor-based parties and their allied unions in Venezuela, Mexico, and Argentina had strong incentives to bargain with each other. Both sides preferred to avoid costly militancy and to be governed by labor-based parties rather than by other parties implementing the reforms. Yet, while some of the unions discussed in this study collaborated with their allied parties in government, others engaged in costly militant activities, and some of them undermined the governance of labor-based administrations. What conditions explain labor loyalty to or betrayal of long-term party allies? Furthermore, the success of labor unions in obtaining their objectives was varied, whether they chose restraint or militancy. Why did government officials grant concessions in some cases and not in others?

INTEREST, INSTITUTIONS, AND COMPARATIVE ANALYSIS

Various bodies of literature provide important insights into these questions. Interest-based theories have focused on classes, factors, and sectors as the unit of analysis. These demand-driven theories derive the policy preferences of different actors from their economic interests, explaining policies as a result of these demands.[9] Some of them focus on the effect of globalization and market-oriented reforms on the economic interest of diverse sectors, such as exposed and protected or public and private.[10] According to these theories, the policy preferences of

[9] Jeffry Frieden, "Invested Interests: The Politics of National Economic Policies in a World of Global Finance," *International Organization* 45 (Autumn 1991); Peter Gourevitch, *Politics in Hard Times: Comparative Responses to International Economic Crises* (Ithaca, N. Y.: Cornell University Press, 1986); and Ronald Rogowski, *Commerce and Coalitions* (Princeton: Princeton University Press, 1989).

[10] For example, Frieden presents a general argument about sector-driven policy preferences and applies it to unions and exchange rate policies. Frieden (Ibid.); and idem, "Labor and Politics of Exchange Rates: The Case of the European System," in Sanford Jacoby, ed., *The Workers of Nations: Industrial Relations in a Global Economy* (New York and London: Oxford University Press, 1995). Peter Swenson focuses on cross-class coalitions based on sector-level preferences with regard to collective bargaining decentralization and state adjustment. Swenson, "Bringing Capital Back In, Or Social Democracy Reconsidered," *World Politics* 43 (July 1991). The Latin American literature uses a similar logic to explain populist coalitions between urban workers and industrialists producing for the domestic market based on the transfer of resources from exporting to protected sectors for import substitution industrialization. Fernando Henrique Cardoso and Enzo Faletto, *Dependencia y desarrollo en América Latina* (Dependency and development in Latin America) (Buenos Aires: Siglo XXI, 1969); and Guillermo O'Donnell, *Modernization and Bureaucratic-Authoritarianism* (Berkeley: University of California, Institute of International Studies, 1973).

employees (and employers) in the internationally exposed (or protected) sectors are based on their economic interests in international markets or their dependence on state subsidies. These theories provide a robust account of the origin of union preferences, based on union interests vis-à-vis economic liberalization. These theories, however, present some empirical limitations for explaining the patterns of interaction in my case studies because I found variation within sectors that share a common economic interest. For instance, although both Mexico and Venezuela are oil exporters and oil production in both is monopolized by the state, their oil workers' unions reacted to management attempts to increase labor productivity very differently.

Interest-based theories overlook the effect of institutions on shaping social demands. Institutional variables influence the bargaining power of unions and their relationship with governments. In particular, explanations based on the economic interest of unions alone cannot account for the influence of populist legacies or of organizations on members who are uncertain about the effect of market-oriented reforms on their future interests. Institutional analyses focus either on "formal-public institutions" that influence the responses of governments to social demands or on "socioeconomic institutions" that shape the distributional pressures from social organizations.[11] Regarding formal-public institutions, the recent literature on the politics of economic liberalization centers on policymakers and looks at the capacity of governments to control social demands that may derail market-oriented reforms.[12] Regarding socioeconomic institutions, the literature on corporatism emphasizes the effect of macrolevel organizational variables, including union internal dynamics and long-term partisan alliances, on union

[11] Geoffrey Garrett and Peter Lange, "Internationalization, Institutions, and Political Change," in Robert Keohane and Helen Milner, eds., *Internationalization and Domestic Politics* (Cambridge: Cambridge University Press, 1996).

[12] Volumes by Joan Nelson, Steven Haggard and Robert Kaufman, and Dani Rodrik assume that liberalization has concentrated costs and diffused benefits, making it difficult to organize collective action in support of reform. Nelson, ed., *Economic Crisis and Policy Choice: The Politics of Adjustment in the Third World* (Princeton: Princeton University Press, 1990); idem, *Fragile Coalitions. The Politics of Economic Adjustment* (New Brunswick, N.J.: Transaction Books, 1989); Haggard and Kaufman, eds., *The Politics of Economic Adjustment* (Princeton: Princeton University Press, 1992); idem, *The Political Economy of Democratic Transitions* (Princeton: Princeton University Press, 1995); and Rodrik "Understanding Economic Reform," *Journal of Economic Literature* 36 (March 1996). Therefore, the management of reform implies the thwarting of societal resistance and the insulation of reformist policymakers. Catherine Conaghan and James Malloy and Carlos Acuña and William Smith also point out the combination of repression, co-optation, and insulation by skillful policymakers for the implementation of market reforms for Latin America. Conaghan and Malloy, *Unsettling Statecraft: Democracy and Neoliberalism in the Central Andes* (Pittsburgh: University of Pittsburgh Press, 1994); Acuña and Smith, "The Political Economy of Structural Adjustment: The Logic of Support and Opposition to Neoliberal Reform," in Smith, Acuña, and Eduardo Gamarra, eds., *Latin American Political Economy in the Age of Neoliberal Reform* (New Brunswick, N.J.: Transaction Publishers, 1994).

preferences and strategies vis-à-vis governments.[13] The West European version of corporatism assumes open economies and societal corporatism. The Latin American version assumes closed economies and puts greater emphasis on the use of state institutions to control labor organization.[14]

While interest-based theories disregard institutions, macrolevel institutional analysis does not sufficiently explain variation in union behavior within the same country.[15] My case studies, though, show diversity in the union-government interaction within each country. Hence, following a path set by Pizzorno and Crouch[16] and reexamined by the recent comparative literature on union politics and the effects of globalization or market reforms on different sectors and across different countries,[17] I shift the focus of analysis to unions as organizations and to their interactions with governments in a variety of contexts. I pro-

[13] A number of scholars have analyzed the effect of union density and concentration, centralization of wage bargaining, and partisan affiliation on union behavior. See, for example, Phillipe Schmitter, "Still the Century of Corporatism," *Review of Politics* 36 (January 1974); David Cameron, "Social Democracy, Labor Quiescence, and the Representation of Economic Interest in Advanced Capitalist Society," and Peter Lange, "Unions, Workers, and Wage Regulation: The Rational Bases of Consent," in John H. Goldthorpe, ed., *Order and Conflict in Contemporary Capitalism* (Oxford: Clarendon Press, 1984); Lars Calmfors and John Driffill, "Centralization and Wage Bargaining," *Economic Policy* 3 (April 1988); R. Michael Alvarez, Geoffrey Garrett, and Peter Lange, "Government Partisanship, Labor Organization, and Macroeconomic Performance," *American Political Science Review* 85 (June 1991); and Peter Lange and George Tsebelis, "Strikes around the World: A Game Theoretic Approach," in Jacoby (fn. 10).

[14] In Schmitter's original definition, organized interests in "societal" corporatism emerged more autonomously from the state than in "state" corporatism. In the comparative Latin American literature, Ruth Berins Collier and David Collier emphasize state "incorporation" of labor, and Francisco Zapata stresses the deeply political character of union activity resulting from the high degree of state intervention. Country studies confirm this view. In particular, these three countries are classified by both Zapata and Collier and Collier as having historically high levels of state intervention and legal benefits for formal workers. Collier and Collier (fn. 3); idem, "Inducement versus Constraints: Disaggregating 'Corporatism,'" *American Political Science Review* 73 (December 1979); Zapata, *El conflicto sindical en América Latina* (Labor conflict in Latin America) (Mexico City: El Colegio de Mexico, 1986); and idem, *Autonomía y subordinación en el sindicalismo latinoamericano* (Autonomy and subordination of Latin American unionism) (Mexico City: Fondo de Cultura Económica, 1993).

[15] Katrina Burgess's dissertation, an institutionalist work on the reshaping of the alliance between organized labor and labor parties, focuses on the external costs created by political parties on the decision making of unions at the national level. Instead, I propose to analyze the internal dynamics of unions and the effect of competition for leadership or for members on their relations with political parties. Burgess, "Alliances under Stress: Economic Reform and Party-Union Relations in Mexico, Spain, and Venezuela" (Ph.D. diss., Princeton University, 1998).

[16] Allesandro Pizzorno, "Political Exchange and Collective Identity," in Colin Crouch and Pizzorno eds., *The Resurgence of Class Conflict in Western Europe since 1968*, vol. 2 (New York: Holmes and Meier Publishers, Inc., 1978); and Crouch, *Trade Unions: The Logic of Collective Action* (Cambridge: Fontana Paperback, 1982).

[17] Volumes edited by Miriam Golden and Jonas Pontusson and by Christopher Candland and Rudra Sil provide a nice sample of new work in this direction for the developed and the developing world respectively. Golden and Pontusson, eds., *Bargaining for Change: Union Politics in North America and Europe* (Ithaca, N.Y.: Cornell University Press, 1992); and Candland and Sil, eds., *Industrial Relations in the Age of Globalization* (Oxford: Oxford University Press, forthcoming).

isan theory of union-government interactions that can be h to national confederations across countries and to individ- in different sectors within the same country by focusing on ion of variables that can be singled out in different contexts. rison across countries and sectors allows me to isolate the ef- est at the sector level and of institutions at the national level.

ATTERNS OF UNION-GOVERNMENT INTERACTION

tion between unions and governments involves either union r restraint and the capacity to obtain concessions from the t through these means. Militancy here means union-orga- st that affects labor relations; it is the most usual measure- nion behavior. Militancy can disrupt production and governance, especially of labor-based parties who are in use part of their electoral appeal is based on their control of ancy is usually measured by counting the number of strikes, on, and their scope. Repertoires of protest vary, however, de- institutional and cultural legacies.[18] Other means of protest nonstrations, boycotts, sabotage, hunger strikes, sit-ins, and

raction does not end with union militancy or restraint be- overnment can respond by granting or refusing concessions. ncy is costly for unions, union leaders prefer to threaten in- ion rather than actually to exercise it. If the union is strong, should suffice to obtain concessions from the government nflict also involves costs for the government. Yet, some ns choose militancy to achieve their demands despite the d, even weak unions that have more to lose and less to win conflict sometimes opt for "heroic defeats," to borrow etaphor.[19]

rstand this apparent irrationality in the behavior of unions s effect (if any) on the government granting them conces- nbined the reactions of both actors into four possible inter-

effect of institutional and cultural constraints, militancy was measured using diverse nce, the meaning of a general strike in Argentina, where they have occurred often in nder Peronist administrations—is different from that of a general strike in Venezuela, e no antecedents of such means of protest for economic demands. See James W. *ism without Peron: Unions, Parties, and Democracy in Argentina* (Stanford, Calif.: Stan- Press, 1997); and Steve Ellner, *Organized Labor in Venezuela, 1958–1991* (Wilmington, Resources, 1993).

lden, *Heroic Defeats: The Politics of Job Loss* (Cambridge: Cambrige University Press,

TABLE 1
CLASSIFICATION OF POSSIBLE UNION-GOVERNMENT INTERACTIONS

	No Militancy	*Militancy*
Concessions	Successful restraint or "cooperation"	Successful militancy or "opposition"
No concessions	Unsuccessful restraint or "subordination"	Unsuccessful militancy or "resistance"

actions. Four categories resulted from the combination: successful militancy or "opposition," unsuccessful militancy or "resistance," successful restraint or "cooperation," and unsuccessful restraint or "subordination." The classification is presented in Table 1.

I applied this classification to the main national labor confederations in Argentina, Mexico, and Venezuela, as well as to unions in five economic sectors deeply affected by market reforms: automobile, education, electricity, oil, and telecommunications.[20] The national labor confederations were the Argentine CGT, the Mexican CTM, and the Venezuelan CTV.[21] The individual unions organized sectors that were under state management, with the exception of the automobile sector, which enjoyed preferential protection in all three countries.[22] These sectors were chosen because their previous conditions strengthened their bargaining power during the period of protectionism but made them more vulnerable to economic liberalization and state retrenchment. This research design facilitates comparisons across countries, sectors, and even different levels of union organization and facilitates the isolation of the common effect of the independent variables on the dependent variables in these diverse contexts. Since about half of the unions experienced a change in strategy during the period under dis-

[20] National confederations are multisectoral economy-wide organizations to which industry-specific unions adhere.

[21] In all three countries, I compare the responses of the main national confederations to a set of policies: stabilization, privatization, trade liberalization, social security, changes in the regulations for labor organization, and labor market flexibility.

[22] The unions involved in the study in Argentina were the Union of Automobile Workers (SMATA), the Federation of Light and Power Workers (FATLyF), the Federation of Telephone Workers (FOETRA), the Union of State-Owned Oil Workers (SUPE), and the Argentine Federation of Teachers (CTERA) with its rival unions. In Mexico, the unions in the study were the Mexican Union of Telephone Workers (STRM), the Mexican Union of Oil Workers (STPRM), the Mexican Union of Electricity Workers (SME), the National Union of Education Workers (SNTE), and the local union of Ford Motors Workers at Cuatitlán. In Venezuela, the unions studied were the Federation of Telephone Workers (FETRATEL), the Federation of Electricity Workers (FETRAELEC), the Federation of Oil Workers (FEDEPETROL), the Ford Motors' Section of the Federation of Automobile Workers (FETRAUTOMOTRIZ), and the multiple unions in the education sector. In all these cases, I analyzed the process of industrial restructuring and reform that involved bargaining with the specific union studied.

TABLE 2
UNION-GOVERNMENT INTERACTIONS BY COUNTRY AND SECTOR

Sectors	Venezuela	Mexico	Argentina	Sectoral Variation	Total Cases
	Countries				
Oil	C[a] → O	S[b] → S	C → C	C=3, O[c]=1, S=2	6
Automobile	C → C	R[d] → S	O → C	C=3, O=1, S=1, R=1	6
Telecom.	C → O	C → C	O → C	C=4, O=2	6
Electricity	C → O	C → O	O → C	C=3, O=3	6
Education	R → R	O → O	R → R	R=4, O=2	6
Nat'l. conf.	O → O	S → S	S → C	C=1, O=2, S=3	6
Nat'l Variation	C=5, O=5, R=2	C=3, O=3, S=5, R=1	C=6, O=3, S=1, R=2		
Total cases	12	12	12		36

[a] C= cooperation (successful restraint)
[b] S= subordination (unsuccessful restraint)
[c] O= opposition (successful militancy)
[d] R= resistance (unsuccessful militancy)

cussion, two observations are reported for each union, thus doubling their number (N=18x2=36) to show the continuity or change in union strategies.[23] Table 2 summarizes all studied interactions by country and sector. The arrows show the passing of time between the first and the second observation. The indicators for militancy and concessions are provided in the Appendix.

Table 2 shows that the case studies provide an array of observations of different union-government interactions with no apparent national or sector-level pattern. It illustrates the need to go beyond country and sector-level variables. The bottom row of the table shows that each of the countries presented at least three different types of interactions despite common national contexts (defined in terms of political institutions and macroeconomic conditions). In turn, the data in the right-hand columns of the table also show diversity in the interactions within each of the economic sectors despite the similarity of interests. In the next section I propose a hypothesis to account for this variation.

A PARTISAN THEORY OF UNION-GOVERNMENT INTERACTION

Union members have different preferences for wages, work conditions, and job stability derived from their diverse labor-market alternatives.

[23] Changes in the independent variable from one category to another were used as cut-off points when possible. When there were no changes in the independent variable, the cut-off was based on two diverse reform initiatives (for example, privatization and restructuring) when possible.

Union leaders thus aggregate a particular combination of such preferences into specific union demands. Union leaders also organize collective action (either strikes or restraint) and make possible the control of workers' behavior for the intertemporal exchange of current restraint for future benefits.[24] Hence, union leaders, as workers' representatives, are the agents of any exchange with government officials.

Unions, however, are not perfect agents, and the preferences of union leaders need not be the same as those of workers.[25] Union leaders also have their own objectives in the exchange besides those of their constituencies. For example, union leaders may seek ideological or material rewards while acting as workers' representatives, or they may prefer to maximize the long-term rather than short-term goals of workers. It is possible, however, to assume that they want to maximize leadership survival. That is, whatever the objectives of union leaders, their primary constraint is to "remain in power because otherwise they would not be able to pursue their objectives."[26] Hence, while their objectives can be perceived as a cost for their agency role in the exchange, union leaders are constrained by their constituencies' preferences to the extent that they want to avoid being replaced as workers' agents.[27] Union leaders want to avoid replacement by internal or external rivals. They can be replaced by new leaders who propose a different set of union demands (partisan or leadership competition), or their members can leave the union for other unions whose banners have become more attractive for them (union competition).[28]

Previous attempts at leadership survival have led union leaders to build long-term alliances with political parties to complement industrial muscle with political influence. Political parties channeled labor demands through the state and helped mobilize support from other sectors.[29] Partisan alliances built on historical exchanges created loyalty

[24] Pizzorno (fn. 16), 278.

[25] Crouch (fn. 16), 161. The imperfection of the union agency may be a desirable goal for workers that select union leaders not just to carry their demands but also to articulate them and to calculate the benefits of intertemporal exchanges.

[26] Henry Farber, "The Analysis of Union Behavior," in O. Ashenfelter and R. Layard, eds., *Handbook of Labor Economics* (Amsterdam: North-Holland, 1986), 1080.

[27] Farber argues that the democratic constraints on the leadership range from cases where the limits are so loose that the leadership can maximize their objective function without regard to the constraints of the political process (dictatorship) to cases where the leadership is severely hampered by the political process and the need to answer the rank and file. Yet, he shows that the possibility of insurgency constrains leaders even in imperfect democracies. Ibid.

[28] In a Hirschmanian sense, replacement by alternative leaders can be assimilated to "voice" within the same organization while the abandonment of the union by members is similar to his concept of "exit." Albert Hirschman, *Exit, Voice, and Loyalty* (Cambridge: Harvard University Press, 1970).

[29] J. Samuel Valenzuela, "Labour Movements and Political Systems: Some Variations," in Marino Regini, ed., *The Future of Labour Movements* (Newbury Park, Calif.: Sage Publications, 1992).

onds between parties and unions, thus influencing both their ability to argain and the attitudes of union leaders toward incumbent politiians. Thus, partisan loyalty implies that, all things being equal, union eaders should be more willing to restrain their militancy when their alied parties are in the government and to increase it when those parties re in the opposition.[30] It follows that if allied union leaders are relaced by others associated with opposition political parties, the change vill affect not only the union leaders but also the terms of unionovernment interactions.

Strategic politicians consider the effect of partisan loyalty on union ttitudes. Loyal unions can facilitate the implementation of governnent policies and provide electoral support. Incumbent politicians hould prefer to reward their loyalty rather than support unions that ave no attachment to the governing party[31]—in particular, nonallied nion leaders, who have no ideological and electoral attachments to the overnment and who, therefore, have fewer incentives for restraint. The dentity of the party in government thus affects the organizational dyamics of labor unions. Partisan loyalty may facilitate fulfilling contituencies' demands at the time of the original alliance. If, however, narket reforms implemented by labor-based parties increase the uncerainty of workers about their future, partisan loyalties can become conradictory with constituencies' demands, thereby affecting leadership urvival and the incentives of labor leaders in their interaction with overnments.

ARTISAN OR LEADERSHIP COMPETITION

Diversity in labor partisan affiliations implies that various parties appeal to organized labor. In Europe these parties tend to be the Comnunist, the Socialist, and the Christian Democratic Parties. In Latin America strong populist labor parties have historically competed with eftist parties for union influence. The higher costs of leadership cometition in a less democratic context also increased the value of subsiies provided by political parties in Latin America. Leadership ompetition between rivals associated with different political parties ould take place within a single union in the form of partisan competiion or across diverse unions combining partisan and union competi-

[30] Walter Korpi explains this pattern as the result of a trade-off between industrial and political reurces. Korpi, *The Working Class in Welfare Capitalism* (London: Routledge and Kegan Paul, 1978). In e Latin American context, Zapata explains the empirical regularity of political strikes by the influce of the state on industrial relations. Zapata (fn. 14, 1986).

[31] The cost of bargaining with nonallied union leaders is higher due to the lack of mutual trust while rt of the agency costs may indirectly feed the coffers of opposition parties.

tion for membership. A partisan monopoly exists when the union leadership affiliated with the labor party faces no competition, such as in the Mexican telephone union. Partisan competition occurs when activists of diverse parties compete or even share leadership positions through proportional representation in the same organization, such as in the Venezuelan CTV.

In some cases, diverse factions of the same party compete for union leadership. Such competition still implies the risk of the allied union leaders being replaced as representatives. For the incumbent labor-based parties, partisan competition implies the threat of allied union leaders being replaced by hostile activists associated with the electoral opposition. Leadership competition among factions will only imply such a threat if the winning faction is likely to defect from the party and join the opposition.

If allied union leaders perceive that leadership competition grows (for example, by losing local union elections), they will try to recover the following of their constituencies to avoid replacement. If they believe that rival leaders are taking advantage of their restraint vis-à-vis market reforms, their incentives for militancy will increase to avoid replacement. Calls for militancy aim to show their constituencies that they have not "sold-out." Leadership survival increases incentives for militancy—even if it is not the best bargaining strategy considering their available information—because if labor leaders are replaced, they will not be able to continue bargaining.

Union Competition

Union competition, or organizational fragmentation, is the rivalry among unions for the representation of workers in the same sector. Where a union monopoly exists, a single union represents all the workers in the sector. In Mexico, for example, the teachers' union was the only union in the education sector, and teachers had no option but to join. Where there is union competition, however, several organizations vie for the membership of the same workers. In Venezuela, for instance, thirteen federations competed for the representation of teachers by 1991.

Union competition introduces the need to coordinate the action of different unions to organize collective action, whether militancy or restraint. The larger the number of unions competing for the same members, the harder it is to coordinate them for collective action.[32] Union competition makes coordination and its enforcement more difficult by

[32] Mancur Olson, *The Logic of Collective Action* (Cambridge: Harvard University Press, 1971), 48.

reating incentives to attract members from rival organizations, thereby ncreasing coordination costs for unions engaging in collective action.[33] Jnions trying to attract members from rival organizations are more ikely to differentiate themselves by breaking coordination to become he most appealing to potential members. This situation is heightened f unions have diverse partisan identities that generate not only differnt attitudes toward the government but also different bargaining costs mong competing unions.

In addition to coordination costs, each union is also weaker than a ingle monopolistic union because each leader controls the militancy or estraint of only a share of the involved workers. As a result, governnent officials put less value on the exchange with individual unions beause they have to enter multiple interactions in their negotiations with he sector, thereby increasing bargaining costs. Thus, although governnent officials prefer to reward loyal unions, they should be less likely to nake concessions to competing unions. Government response to loyal nions, therefore, is more related to the strength of the union than to ts militancy.

HE PARTISAN THEORY AND UNION-GOVERNMENT INTERACTIONS

he effect of union competition and leadership competition on the ineraction between union leaders and government officials varies acording to the party in power. When labor-based parties implement narket-oriented reforms, allied union leaders are willing to collaborate espite the uncertainty and distress of their constituencies due to their oyalty to long-term allies. Loyal union leaders are predisposed to colborate and can expect some concessions in return. Moreover, they ave better information about the commitments and constraints of heir partisan allies and the need to implement these policies.

Yet, when labor-based parties shift their policy, they leave void the opulist space of those who disagree with market-oriented policies. Opposition political parties from the left or disgruntled politicians plitting from the governing party (on ideological grounds) can take dvantage of the policy shift to occupy that space. This movement may acilitate the growth of militant union activists allied with political paries or partisan factions opposed to market-oriented reforms and may

[33] Miriam Golden shows that coordination in wage bargaining is most likely when union monoply is high because the competition for members "provides a strong incentive for unions to try to maxnize their wage gains in order to retain members or to attract them away from competitors." Golden, The Dynamics of Trade Unionism and National Economic Performance," *American Political Science Review* 87 (June 1993), 441.

enhance leadership and partisan competition. Indeed, in Mexico, Venezuela, and Argentina, new political parties occupying this populist space emerged.[34] Furthermore, this movement more likely turns factional competition into partisan competition. Partisan competition, in turn, increases the likelihood of militancy. These effects are more likely in sectors where the high uncertainty of workers about the sudden shift from public to private or from protection to exposure is likely to prompt them into militancy. Additionally, market-oriented reforms can sharpen union competition by provoking splits over how to respond or by aggravating the distributive struggle for shrinking resources among rival unions. In their attempts to attract members from rival organizations to increase their representation in these disputes, unions have more incentives to boycott coordination efforts as a strategy of differentiation, thereby weakening all the unions within the sector.

These explanatory conditions are not fixed. The party in power may change with elections. Union competition and partisan competition within the union movement can also change. Governments, though, cannot usually manipulate these variables in the short-term, although their policies can influence their changes. Market reforms in particular can affect all these variables by improving or damaging the electoral opportunities of incumbents, thus affecting the likelihood of challenges to labor leaders and making it harder for unions to coordinate their strategies. If these changes in the explanatory conditions occur, changes in the dependent variable should follow, thereby explaining variations in the union-government interaction during the period under discussion. Table 3 summarizes the expected union-government interactions when a labor-based party implements market-oriented reforms.

"Cooperation," or successful restraint for concessions, is more likely in the absence of union and partisan competition—when only one union organizes all workers and is affiliated with the governing party. Partisan loyalty reduces the incentives for militancy and facilitates bargaining while union monopoly boosts the bargaining power of the union because government officials want the collaboration of a strong and loyal union.

"Opposition," or successful militancy, can more likely be expected in the presence of growing partisan competition and union monopoly—

[34] Market reforms created a new critical juncture in the electoral politics of these countries (together with a simultaneous process of democratization or decentralization) that resulted in the emergence of new parties or divisions in the incumbent labor-based parties. Argentina experienced the emergence of FREPASO or Front for a Country with Solidarity (originating in a splintering of Peronism). Mexico saw the organization of the PRD or Party of the Democratic Revolution (also emerging from a division within the PRI). In Venezuela, Causa R grew to become a national political party and was followed by an array of new political options that reshaped the traditional two-party system.

TABLE 3
PREDICTED BACKGROUND CONDITIONS FOR LOYAL UNIONS AND LABOR GOVERNMENTS

		Partisan Competition for Leadership	
		Monopoly (one party)	*Competition (many parties)*
Union Competition for Members	Monopoly (one union)	Cooperation (successful restraint)	Opposition (successful militancy)
	Competition (many unions)	Subordination (unsuccessful restraint)	Resistance (unsuccessful militancy)

when leaders affiliated with different parties compete for the control of a single union. Growing partisan competition based on protesting the consequences of the policy shift increases the incentives for "irrational" militancy[35] because allied union leaders are afraid of being replaced and the sections controlled by contenders have already turned militant. Since the union is strong, government officials are more likely to grant concessions so that allied union leaders can show a better record than their contenders.

"Subordination," or unsuccessful restraint, more likely results from competition among different unions affiliated with the governing party. While partisan loyalty facilitates restraint, union competition weakens all unions despite their loyalty. Government officials may also choose to reward only the most compliant unions. This selection should also prompt competing unions to comply to avoid losing resources and becoming less attractive for members than other competing unions.

"Resistance," or unsuccessful militancy, more likely happens when rtisan competition and union competition overlap—competing ions affiliated with different parties. Union competition weakens all ions and, together with partisan competition, makes coordination ›re difficult. Unions associated with opposition parties protest to dif-entiate themselves from cooperative loyal unions. If they succeed in racting members due to their bellicose stance, loyal unions will turn ›re militant to avoid losing members although union competition ıkes them weak and unlikely to be successful.

' Since union monopoly makes the union strong and the government can observe this strength, it ns unnecessary to enter a conflict to show it in terms of the external interaction, but partisan com-tion prompts union leaders to protest for reasons linked to their internal power.

APPLYING THE PARTISAN THEORY IN ARGENTINA, MEXICO, AND VENEZUELA

According to the partisan theory, union-government interactions in Argentina, Mexico, and Venezuela should have felt the effects of partisan ties because the incumbent parties were labor-based. In this context, leadership competition should have increased the likelihood of labor militancy, and union competition should have decreased the likelihood of concessions for unions. This section provides a brief description of the dynamics of the case studies, relating them to the predicted background conditions for each interaction, which are summarized in the Appendix.

In Venezuela, although AD historically controlled the CTV, other parties competed for leadership and displaced AD briefly in the 1960s. Also the system of proportional representation reinforced the pluralism of the CTV in its leadership, resulting in the inclusion of minority parties in the executive committee.[36] In 1989, after Pérez announced the reforms, urban riots signaled the general discontent of the population. Union leaders of AD, afraid of losing control of the CTV to opposition parties protesting the reforms, called the first economic general strike in Venezuelan history for May 1989. The tension between partisan loyalty and leadership survival divided the AD union leadership between those with positions in the CTV who were more favorable to the strike and those with party appointments who were more reluctant about it.[37] Pérez's concessions included emergency wage hikes, suspension of layoffs, and retaining price controls for basic staples. In 1990 and 1991, however, Causa R, a new political party, grew rapidly in the union movement by rejecting reforms, especially among public sector workers, further increasing partisan competition. As a result, the CTV called additional protests despite Perez's move to halt the reform of the severance payment system and social security and his acceptance of union demands for the resignation of the labor minister. Hence, union monopoly and increasing partisan competition explain the CTV opposition.

In spite of the national tendencies in union behavior, Venezuela experienced variation in union-government interactions. In the case of the privatization of the state-owned monopoly of telecommunications,

[36] Ellner (fn. 18).

[37] Author's interviews with AD union leaders of different factions confirmed the account of Steve Ellner. See Ellner, "Organized Labor's Political Influence and Party Ties in Venezuela: *Acción Democrática* and Its Labor Leadership," *Journal of Interamerican Studies and World Affairs* 31 (Winter 1989).

the combination of union monopoly and partisan monopoly exercised by the AD union leadership resulted in cooperation. Labor restraint was exchanged for concessions that included employee-owned stock, job stability, and labor directors. Subsequent worker discontent with the privatization, however, resulted in Causa R winning the union elections in the main local union (Caracas) while growing in other regions by protesting market reforms and privatization. As a result, a new AD leadership took control of the union and increased the militancy of the national union against labor restructuring to show their responsiveness to the discontented rank and file. In spite of attempts at broad restructuring, the company could not change work conditions from the privatization agreements. In this case, the increase in partisan competition combined with union monopoly to move the union into opposition.[38]

Very similar dynamics explain the shift from cooperation to opposition in the cases of the electricity workers' union of the state-owned electric company (CADAFE) and of the state-owned oil company (PDVSA). Union monopoly and an unchallenged AD leadership facilitated cooperation when the Pérez administration decentralized CADAFE. Concessions to the union included handsome monetary incentives for transfers, job stability, and wage increases for workers in the interior. Yet, the subsequent discontent of workers also gave the election of the Caracas's union to Causa R, who had led the local union into a wildcat strike. Subsequently, the AD leadership increased the militancy of the union and successfully boycotted industrial restructuring. Again, the increase in partisan competition combined with union monopoly resulted in opposition. In PDVSA, the combination of union monopoly and no partisan competition for the AD leadership favored cooperation in 1991. In return, the union retained hiring prerogatives. Afterward, workers' discontent resulted in increasing partisan competition, not only from Causa R but also from other left-wing parties—the Movement to Socialism (MAS) and the Electoral Movement of the People (MEP). As a result, the AD leadership took a more combative stance against management proposals in 1993 and tried to gain legitimacy by decreasing its discretionary use of hiring prerogatives and obtaining a halt to the reform of commissaries.[39] Thus, partisan competition, along with union monopoly, explains the shift of the union to opposition.

Contrary to the unions in public enterprises, the two remaining cases show a different pattern. The union of Ford Motors, a local union

[38] Union leaders, government officials, and company managers, interviewed by the author, Caracas, June–July 1994, May–July 1996.
[39] Ibid.

based in Carabobo, opted for cooperation. The union restrained and accepted layoffs and new working rules in return for participation in the selection of those laid off, reincorporated, and trained. In this case, cooperation resulted from the combination of union monopoly and partisan monopoly. In a personal interview, the labor relations' manager said that cooperation for restructuring and training was possible because the AD union leadership did not face competition from left-wing parties.[40] In education, the overlap of union competition and partisan competition in thirteen federations affiliated with diverse political parties and groups resulted in militant but unsuccessful resistance to the administrative reform initiated by Minister Gustavo Roosen. While coordination was difficult, the militancy of the rank and file rewarded the free riding of bellicose organizations. This situation brought the vice minister of education, Francisco Castillo to complain, "I think that there is a competition among unions, where the one that strikes the most is the one that is the most supported by the rank and file."[41] In spite of the differences between these two cases, both sectors contrast with telecommunication, electricity, and oil in that the levels of union competition and partisan competition did not change during the period studied, and neither did the studied interactions.

In Argentina,[42] Peronism's policy shift divided the CGT. All three competing factions, though, remained within the ranks of Peronism. The Menem administration was unwilling to make concessions, in particular to the most "populist" faction, although it provided selective incentives (such as guaranteed monopolies of representation and execu-

[40] Author's interview with Ford's labor relations manager and AD union leader was confirmed by Consuelo Iranzo, Luisa Bethencourt, Hector Lucena, and Fausto Sandoval Bauza. See Iranzo, Bethencourt, Lucena, and Bauza, "Competitividad, Calificación y Trabajo: Sector Automotriz Venezolano" (Competition, qualifications, and work: Automobile sector in Venezuela) (Manuscript, Cendes, 1996).

[41] Cite from *El Nacional*, January 9, 1991. Former Minister Gustavo Roosen and union leaders of FetraMagisterio and Fenatev confirmed this account derived from a press chronology, in interviews with the author, Caracas, June 1996.

[42] The information on the Argentine cases is derived from a press chronology, union annual reports and other documents, interviews with union leaders of all factions, three ministers of labor and other government officials, as well as labor relations managers in the involved companies. The Argentine press chronology was elaborated in the archives of the newspaper *Clarin* and includes newspapers such as *Clarin*, *La Nacion*, *La Rozon*, *Cronica*, *ElCronista Comercial*, *Pagina 12*, and *Ambito Financiero*. Union sources included the internal constitution, annual minutes, and balances for the 1988 to 1994 period for SMATA, FATLyF, CTERA, FOETRA, and SUPE. Documents included collective bargaining contracts signed by the unions and approved by the Ministry of Labor, bill proposals, parliamentary information on introduced, modified, and passed bills. Interviews included SUPE and CGT union leader Antonio Cassia (Buenos Aires, 1993, 1995), FATLyF union leader Carlos Alderete (Buenos Aires, 1993), CTA union leader Victor De Gennaro (Buenos Aires, 1995), CTERA union leader Marta Marffei (Buenos Aires, 1995), ministers of labor Armando Caro Figueroa (Buenos Aires, 1994), Rodolfo Diaz (Buenos Aires, 1992, 1995), and Enrique Rodriguez (Buenos Aires, 1992, 1995), managers of labor relations, such as YPF's Roberto Teglia (Buenos Aires, 1995), and Telecom's Juan Giar (Buenos Aires, 1995).

tive appointments) to the most compliant. Peronist labor unions were unable to halt a new law that introduced temporary hiring and decrees that established wage restraint and deprived them of collecting and administering union health-fund fees. As a result, many important unions moved out of the populist faction, and in 1992 all three factions decided to unify under a leadership dominated by the most cooperative union leaders. Thus, from 1989 to 1992 the combination of partisan monopoly and union competition subordinated Peronist unions to the administration. After its unification, the CGT recovered its union monopoly and maintained its Peronist loyalty, moving from subordination to cooperation. As a result, it was rewarded with concessions that included changes in the reforms of pensions, social security, subsidies for health funds, and the maintenance of legislation on collective bargaining and labor organization.

Argentina also shows diversity in union-government interactions although national institutions and conditions remain the same. Menem also privatized the state-owned monopoly of telecommunications, ENTEL. Unlike its Venezuelan counterpart, the union opposed privatization by increasing its militancy. A combination of union monopoly and leadership competition explains this outcome. Leadership competition in the telephone workers' union increased when a populist faction that opposed market reforms and privatization won control of the union in Buenos Aires (the largest in the federation) and increased union militancy.[43] A monopolistic union, nonetheless, was able to obtain concessions, including the administration and representation of employee-owned stock, handsome retirement packages, and subsidies for the union health fund. After privatization, the loyal Peronist leadership won control of the rebellious locals and curtailed leadership competition. Hence, with the absence of union and partisan competition, the union cooperated with the new private management in discussions of labor restructuring. Concessions at this stage included union participation in training, contracts for union-organized companies, and voluntary retirement packages. The privatization of state-owned electric companies also resulted in initial opposition due to a combination of union monopoly and leadership competition. The local unions, which the populist militant faction controlled and which competed with the

[43] The policy-shift of Peronism encouraged the formation of a splinter group, which together with some left-wing parties, formed, in turn, a new opposition political party in 1992 that would became the FREPASO in 1994. Among its core organizers was a group of union leaders that had broken ties with the Peronist union movement and organized a new small confederation, the CTA (Congress of Argentine Workers) in 1992.

loyal leadership, joined the CTA in 1992. Since the national federation was a member of the CGT, the dissident unions were expelled. As a result, leadership competition receded. Lacking union competition and partisan competition, the union began to cooperate. In the first period, union concessions included employee-owned stock, subsidies for the union health fund, and contracts for union-organized cooperatives of former employees. In the second period, they included subsidies for union participation in the privatization of companies—public utilities, an energy transmission company, and the concession of a coal mine—as well as voluntary retirement packages.

The cases of the restructuring and privatization of the state-owned oil company (YPF) and of the decentralization of education differed from the previous two examples. In the case of the YPF, the Peronist union leadership faced no competition and had a very close relationship with Menem that facilitated bargaining for concessions, which included subsidies for the union to buy the oil fleet, contracts for union companies employing laid-off workers, and voluntary retirement packages. Thus, leadership monopoly and union monopoly resulted in cooperation. Unlike the oil company, the education sector, as in Venezuela, was fragmented into various unions. Many of them gathered into a national confederation, CTERA, whose leader, Mary Sánchez, was one of the Peronist union leaders who had left the party and the CGT to found the FREPASO and the CTA in protest of the policy shift. Yet, many others unions did not belong to the national confederation and competed with CTERA unions in every province. Many of the non-CTERA unions were Peronist. In this case, the overlapping of union competition and partisan competition coincided with resistance or unsuccessful militancy against the decentralization of education to the provincial level enacted by Congress in 1992. In spite of the restraint of oil workers and of the militancy of teachers that accounted for more than a third of total strikes in 1991 and 1992, in both cases the levels of union competition, partisan competition, and union-government interactions remained the same during the period under discussion.

Lastly, the national union of automobile workers moved from opposition to cooperation in 1991. In this case, union monopoly with no partisan competition made the union's initial opposition unexpected, although the subsequent cooperation coincides with the theory. The increasing militancy of the union until 1991 is related to the rejection of trade liberalization in a heavily protected sector. The consequence of this militancy together with lobbying by business was to persuade the government to grant an industrial regime of protection and gradual

opening for the sector after 1991. The industry was thus exceptionally successful in shielding itself from the conditions created by a sudden trade opening. After the regime was granted, the government became again an important partner, and partisan loyalties moved the union toward cooperation.

The Mexican cases highlight the importance of contextualizing the explanatory variables within the constraints created by the political regime on the means of expressing militancy and the costs of partisan competition.[44] Yet at the same time, even in the most constrained environment, where the regime is not totally democratic, the interaction between unions and the government varies. This variation cannot be explained by traditional theories based only on the features of the Mexican regime. The Mexican CTM subordinated to Salinas's policies. The only important concession that it was able to obtain was the prevention of labor-code reform. The same variables—partisan monopoly combined with union competition—explain the confederation's subordination and its exception to cooperation in relation to the labor code.[45] The CTM competed with other national labor confederations also associated with the monopolistic PRI. While partisan loyalty reduced their incentives for militancy, government officials manipulated union competition for scarce resources among the various PRI-related national confederations and rewarded the most compliant of them with selective incentives (such as public recognition and favorable treatment in arbitration boards) to weaken CTM claims. The only exception was when the CTM unified with all other PRI-confederations and reduced union competition to boycott the reform of the labor code successfully. The same institutional mechanism that provided a common forum for all of them to unify, an umbrella organization called the Congress of Labor, had previously facilitated union competition because decision making was based on consensus or unanimity.[46]

[44] State imposed limits on strike activity discard the usefulness of using strikes as a measure of militancy but do not imply that militancy does not take place. Mexican workers have held illegal strikes, used strike petitions as a threat to show their militancy, and resorted very often to sit-ins, demonstrations, and even extreme measures such as going naked or on hunger strikes. On repertoires of protest, see the account of Maria Lorena Cook, *Organizing Dissent* (University Park: Pennsylvania State University Press, 1996). For a historial analysis of the evolution of union strategies, see Alberto Aziz Nassif, *El estado mexicano y la CTM* (The Mexican state and the CTM) (Mexico City: Ed. La Casa Chata, 1989); Ilán Bizberg, *Estado y sindicalismo en México* (State and unionism in Mexico) (Mexico City: El Colegio de México, 1990); and Kevin Middlebrook, *The Paradox of Revolution* (Baltimore: John Hopkins University Press, 1995).

[45] Graciela Bensusán confirms interviews with union leaders and government officials. Bensusán, "Los determinantes institucionales de la flexibilización laboral" (Institutional influences on labor flexibility), *Revista Mexicana de Sociología* 1 (1994).

[46] For instance, the Revolutionary Confederation of Workers and Peasants (CROC) and the Revolutionary Confederation of Mexican Workers (CROM) explicitly boycotted CTM protest against wage ceil-

As in Venezuela and Argentina, the state-owned monopoly of telecommunications, Telmex, was privatized in Mexico. In this case, cooperation in relation to both privatization and restructuring corresponds to the overlap of union monopoly and leadership monopoly, which remained unchanged during the entire period studied. The latter, held since 1974 by a charismatic PRI union leader, Francisco Hernández Juarez, who had co-opted the opposition, successfully provided benefits for his constituencies and developed a close relationship with the president, who often used this union as an example of modern unionism. Moreover, since this union was not affiliated with the CTM, it had more room to maneuver and even organized the Federation of Goods and Services Workers (FESEBES), a new labor confederation competing with the CTM.[47] The union obtained employee-owned shares and a labor director in the private company, wage increases, job stability, and the maintenance of work conditions during privatization. Afterward, concessions included participation in joint committees with management for training and restructuring and increases in wage and nonwage benefits.

The union of electricity workers of the Company of Light and Power, which was not a CTM-affiliate either, also joined the FESEBES. Its charismatic union leader, Jorge Sanchez, also had a close relationship with the president. In this case, the absence of union and leadership competition led to cooperation between the union and management. Cooperation led the government to bail out the company, establish guarantees for union survival, and develop joint union-management committees to discuss productivity and financial issues. Unlike Hernández Juarez, though, the loyal leader lost the union elections in 1993 and was replaced by an independent leadership, breaking the loyalty strings. As expected, this situation increased union incentives for

---

ings in the Congress of Labor and were publicly rewarded by the government. Subsequent exit of unions from the CTM into preferred organizations, most notably the CROC, increased the pressure on its leadership. See my press chronology, which derives from the archives of Entorno Laboral and which includes newspapers such as *La Jornada*, *Excelsior*, *El Sol de Mexico*, *Uno mas Uno*, and *Reforma*.

[47] Ruth Berins Collier and James Samstad analyze the development of the FESEBES and the "new unionism" in "Mexican Labor and Structural Reform: New Unionism or Old Stalemate?" in Riordan Roett, ed., *The Challenge of Institutional Reform in Mexico* (Boulder, Colo.: Lynne Rienner, 1995). Interviews with union leaders, government officials, and company managers confirmed the numerous accounts on the strategy of this union. For telecommunications, see Enrique De La Garza, "Sindicato y restructuración productiva en México" (Union and labor restructuring in Mexico), *Revista Mexicana de Sociologia* 1 (1994); and De La Garza and Javier Melgoza, "Reestructuración y cambio en las relaciones laborales en la telefonía mexicana" (Restructuring and change in labor relations in the Mexican telephone industry), in Jorge Walter and Cecilia Senen González, eds., *La privatización de las telecomunicaciones en América Latina (Privatization of telecommunications in Latin America)* (Buenos Aires: Eudeba, 1998). For electricity, see Javier Melgoza, "El SME y la productividad: Los saldos de la negociación" (SME and productivity: Outcomes of bargaining), *Polis* 93 (1994).

militancy (although less than if the leadership was associated with an opposition party). It also, however, reduced the partisan incentives of the government to grant concessions to the union, which nonetheless included a reduction of productivity targets and the maintenance of job stability. In this case, the tension between partisan loyalty and leadership survival was broken by the replacement of the loyal leader and his succession by an independent and more militant union leadership.

In Mexico, the education sector was not as fragmented as in Argentina and Venezuela. The National Union of Education Workers (SNTE) which also was not a CTM-affiliate, had a union monopoly and was controlled by PRI union leaders. A discontented rank and file, organized by dissidents in the National Coordinating Committee of Education Workers (CNTE), had toppled the previous PRI leadership. The new PRI leader, Elba Ester Gordillo, therefore, had to win legitimacy to avoid replacement. As a result, the SNTE became more bellicose after the change in leadership while including the dissidents within the executive committee through proportional representation and increasing internal debate.[48] In 1991 when the decentralization law was discussed, a leak from the secretary of education mentioned the division of the centralized union. The SNTE not only became more militant but also joined forces with the dissident CNTE. The government responded by granting centralized national work conditions and earmarked budgets for the states as well as salary hikes and nonwage benefits for the union. Hence, the combination of union monopoly and partisan competition resulted in opposition.

The unions of oil workers and Ford automobile workers in Mexico, both CTM-affiliates, provide interesting cases to illuminate the limits of my explanatory framework. In such cases, partisan competition may be punished by a nondemocratic government that can resort to coercion. Both unions faced sharp restructuring due to the opening of their sectors to private capital and international competition. Both attempted to resist the changes and were forced into subordination by the repression of the CTM and the state. The workers of the Ford Motors plant at Cuatitlán rejected restructuring and chose an independent leadership (linked with left-wing parties), which broke the tension between partisan loyalty and leadership survival and brought the local union to resist industrial restructuring—albeit unsuccessfully—in 1988 and 1992. The

[48] Although my account derives from a press chronology and interviews with union leaders, government officials, and experts, the process of modernization and democratization of the union has been widely studied. See, for instance, Cook (fn. 44); and Joe Foewaker, *Popular Mobilization in Mexico, the Teachers' Movement 1977–87* (Cambridge: Cambridge University Press, 1993).

local union not only increased its militancy but also attempted to withdraw from the CTM-affiliated national union. The CTM, upheld by the government and the company, responded to the militant leadership and their supporters with violent repression, forcing the local and the national union into subordination.[49] In the case of the oil workers' union, the PRI union leader had supported the opposition presidential candidate, Cuahutemoc Cárdenas,[50] by delivering votes in the oil regions, although PRI local candidates carried the elections. As a result of the partisan threat, the state gave a fatal blow to the union by putting the union leader in prison under fabricated murder charges and by bringing the union into subordination, thereby shifting the issue of leadership survival from workers to state officials.[51] In both cases, the restrictions on political pluralism of the regime limited partisan competition in the unions and tipped the balance in favor of partisan loyalty by dramatically raising the costs of noncompliance with government officials.[52] The Mexican political regime restricted partisan competition in the union movement, thus limiting the explanatory power of this variable for certain unions. Yet, other non-CTM unions, such as the teachers', telephone workers', and electricity workers' unions—and even the CTM itself—retained their relative autonomy, thus permitting the use of the explanatory variables based on union dynamics to illustrate their interaction with the government.

Table 4 shows the high correspondence between the patterns of union-government interaction presented in Table 2 (and summarized in the Appendix) and the explanatory conditions defined by my partisan theory and presented in Table 3.

This high correspondence between the observed outcomes and the explanatory conditions provides a better account than either macrolevel or sector-level theories for the interactions studied. In a context of al-

[49] Interviews and press releases are confirmed by Marisa Von Bulow, "Reestructuración productiva y estrategias sindicales. El caso de la Ford en Cuahutitlán 1987–1993" (Production restructuring and union strategy: The case of Ford in Cuahutitlán 1987–1993) (M.A. thesis, Facultad Latinoamericana de Ciencias Sociales-Sede México, 1994); and Jorge Carrillo, "La Ford en México: Restructuración industrial y cambio en las relaciones sociales" (Ph.D. dissertation, El Colegio de México, Centro de Estudios Sociológicos, 1993).

[50] Cárdenas split from the PRI because he rejected the policy shift and failed to be selected as its presidential nominee. He launched a new coalition, later to be renamed as the PRD, and he delivered the strongest electoral blow to the PRI by dramatically reducing its share of votes (and even claiming victory stolen by fraud) in 1988.

[51] Fabio Barbosa gives a graphic account of the process that confirms interviews and press releases. Barbosa, "La reestructuración de Pemex" (The restructuring of Pemex), *El Cotidiano* 46 (March–April 1992).

[52] These two cases related to Burgess's argument about the PRI imposing external costs on union leaders' decisions. Burgess (fn. 15).

TABLE 4
EXPLANATORY CONDITIONS AND UNION-GOVERNMENT INTERACTION IN ARGENTINA, MEXICO, AND VENEZUELA

'ossible Interactions	*Explanatory Conditions Assuming Partisan Loyalty*		*Frequency of Each Interaction in the Study*	*Interactions Corresponding with Explanatory Conditions*
	Partisan Competition	*Union Competition*		
'ooperation (successful restraint)	No	No	13	13
;ubordination (unsuccessful restraint)	No	Yes	6	3
)pposition (successful militancy)	Yes	No	12	11
lesistance (unsuccessful militancy)	Yes	Yes	5	5
'otals			36	32

ied labor parties implementing market reforms, partisan competition ınd union competition influenced the interaction between loyal union eaders and government officials in different sectoral and national conexts. Additionally the explanatory conditions in the studied interacions varied as expected over the short-term, further reinforcing the mplications of this explanatory logic.

Other patterns of interaction also deserve attention. In Argentina a)attern of cooperation between unions and government tended to arise nainly from opposition but also from subordination at the national evel. In Venezuela a pattern of opposition between the two arose nainly from cooperation, although arriving early in this interaction at he national level. Finally, although Mexico shows a less clear trend, it s where most cases of subordination concentrate. These trends in the nteractions studied are related to national institutions that can faciliate the emergence of union competition and partisan competition, :hus explaining these patterns.

In Argentina national regulation established monopolies of repreentation for collective bargaining except in the public administration. As a result, union competition is limited except at the peak level and in the public administration, where the cases of union competition leading to subordination or resistance emerged. The strong partisan identity of labor unions combined with their ability to obtain concessions from the government helped Peronist labor leaders retain control of their unions. In Venezuela collective bargaining rules also facilitated effective union

monopolies.[53] The education sector was an exception, excluded from collective bargaining and faced with union competition, as in Argentina. The growth of leadership competition that increased labor incentives for militancy in many AD unions is related to the political changes brought about by the combination of market reforms and political decentralization. These changes made room for new political options not only in the union movement but also in the electoral arena.[54] In Mexico the characteristics of the regime allowed the government to overcome the effect of the explanatory conditions in the cases of oil workers and the Ford Motors' workers of Cuatitlán. These cases do not contradict the explanatory logic but highlight its limitation. In the oil workers' union, partisan monopoly and union monopoly had previously led to labor restraint in return for sizable concessions including nonwage benefits for workers, hiring prerogatives, contracts for union companies, and fees over suppliers' contracts for the union. After its leader challenged the election of Salinas, however, the union was forced into subordination. In the case of Ford Motors, when alternative leaders won local elections, breaking loyalty ties and increasing labor militancy, the government and the CTM also forced the union into subordination. Thus, in Mexico the opportunities for leadership or partisan competition were limited in most cases by a regime that had also curtailed political pluralism in the electoral arena.

National patterns also had important policy consequences. The prevalence of opposition in Venezuela weakened the pace of market reforms until they finally were suspended. The predominance of cooperation in Argentina facilitated the rapid implementation of market reforms and the adoption of union strategies more akin to a market economy. In Mexico, although subordination eased the implementation of reforms, it also slowed the adaptation of union strategies to the new environment. In spite of national trends, though, some cases in each country entered into different interactions explained by other combinations of union competition and partisan competition.

At the sector level there were less clear patterns. The cases in the oil and automobile sectors exhibited a large variation in interactions. Telecommunications and electricity do not have a single predominant interaction, but unions in both sectors successfully obtained concessions through cooperation or opposition (and enjoyed union monop-

[53] International Labour Organization, *Relaciones de Trabajo en Venezuela* (Labor relations in Venezuela) (ILO report, 1991).

[54] Not only did Causa R grow in the unions and in the 1994 elections, but also many political outsiders, including current President Chávez, challenged the traditional political parties after that election.

). In the education sector, both resistance and opposition interac-
ıs are associated with higher militancy. This finding is consistent
h predictions of higher militancy in public-sector workers[55] and with
politicization of the education sector[56] that facilitates the emer-
ıce of partisan competition. While partisan competition in Ar-
ıtina and Venezuela had been an enduring feature of teachers'
anizations, in Mexico it took an internal rebellion to allow for in-
ısing partisan pluralism within the union.[57] Thus, although some
tor-level patterns emerged, they are insufficient to account for the
lied interaction.

CONCLUSIONS

is article attempts to complement the focus on policymakers that
predominated the literature on the political economy of market re-
ns in developing countries by providing some perspective on labor
anizations, which have been neglected. The findings underscore the
ct of partisan identities on creating a tension for union leaders when
ir allies implement market-oriented reforms. In my case studies,
on leaders, on the one hand, were pulled by their partisan loyalties—
ed on long-term goals or on their own personal gain. On the other,
y had to respond to their constituencies—based on the short-term
l of political survival. Partisan loyalty to governing allies on the part
ınion leaders increases incentives for restraint, while leadership or
tisan competition pushes them toward militancy if demanded by
ıstituencies. At the same time, union competition affects the
ngth of the union and its ability to obtain concessions from the gov-
ment. This theory explained most of the interactions in my study
ween union and labor-based governments implementing market-
nted reforms in Argentina, Mexico, and Venezuela. In these coun-
s, partisan loyalties inherited from the original postwar alliance
ween the Peronism, PRI, and AD with labor unions had an important
uence on the transition to open economies. The interaction of par-
n loyalties with union competition and partisan competition, how-

Geoffrey Garrett and Christopher Way, "The Rise of Public Sector Unions, Corporatism and
oeconomic Performance, 1970–1990," in Barry Eichengreen and Jeffry Frieden, eds., *The Politi-
onomy of European Integration* (New York: Springer-Verlag, 1995).
Ivan Nuñez, "Sindicatos de maestros, Estado y Políticas Educacionales en América Latina"
hers unions, state and education policy in Latin America), in M. L. P. B. Franco and D. M. L.
, eds., *Final do Seculo: Desafios da educacao na América Latina (End of the century: Challenges for ed-
n in Latin America)* (Sao Paulo: Cortes Editora, CLACSO REDUC, 1990).
Cook (fn. 44).

ever, can also influence the attitudes of unions and governments toward each other for nonlabor parties in government. The absence of partisan loyalties can make restraint and negotiations more difficult for unions and governments affecting the influence of leadership competition for militant labor leaders and of union competition in noncooperative unions. Further testing in other cases will show the extensions and limitations of this partisan theory.

If my argument is correct and political actors were aware of the effect of union competition and leadership competition, unions should have tried to affect these variables. In the previous section, I analyze the effect of legal institutions and electoral dynamics in shaping national patterns of union competition and leadership competition and thus union-government interactions. According to Collier and Collier the incorporation of labor shaped both legal institutions and the political system.[58] When the alliance between labor unions and political parties was established, the corporatist labor legislation that emerged shaped the opportunities for union competition and leadership competition because unions sought to strengthen their bargaining power and politicians wanted to reinforce labor loyalties. In Argentina, Mexico, and Venezuela, the labor legislation included restrictions to exit (ranging from permission to close shops to monopolies of representation) and facilitated the controls over leadership selection by incumbents.[59] At the same time, the partisan loyalties derived from the alliance were reinforced by the delivery of wage and social benefits when labor-based parties were in government.[60] National patterns thus resulted from the diverse emphasis on these variables that emerged from the original alliance. Union competition, however, was "stickier" than partisan competition because it was harder to modify on an individual basis without a legislative reform. This stickiness explains why leadership competition varied more frequently than union competition in the cases studied, making changes in militancy levels more likely than changes in the ability of unions to obtain concessions in the short term.

Although market reforms affect both conditions as analyzed above, changes are not easy to manipulate in the short term. Politicians cannot easily control electoral dynamics unless they resort to repression. Indeed, they are usually subject to the electoral dynamics unleashed by their shift toward the market and by the new alliances that this critical

[58] Collier and Collier (fn. 2).

[59] Collier and Collier (fn. 14, 1979).

[60] Zapata (fn. 14, 1986).

juncture allowed in the union movement.[61] They could reform the rules for labor organization, but these regulations have remained unchanged in all three countries during the period studied. Reforms to the regulations on labor organization affecting union competition would have also risked changing the partisan ties between labor unions and governing politicians on which incumbents counted to implement their policies and remain in power.[62]

The conclusions of this study resonate with Michels's claim that although organization is the weapon of the weak in their struggle with the strong, "it is the organization which gives birth to the dominion of the elected over the electors, of the mandataries over the mandators, of the delegates over the delegators."[63] He considers, however, that leadership competition should limit the iron rule of oligarchy in organizations because the "ascent of the new leaders always involves the danger, for those who are already in possession of power, that they will be forced to surrender their places to the newcomers. The old leader must therefore keep himself in permanent touch with the opinions and feelings of the masses to which he owes his position."[64] Coincidentally with Michels and Hirshman, this study shows that although union competition provides exit options for workers, it weakens unions. Instead, leadership or partisan competition provides voice options to increase representation but without weakening their bargaining power to the same extent.[65]

Finally, this study suggests that future research should address the interaction between different levels of analysis. Such research would not only look to recover forgotten actors (such as individual unions) from oblivion but also to move beyond the national-level bias to include units of analysis defined at the level of sectors, subnational units, and even organizations. The multilevel research design of this study combined cross-country and within-country comparisons to facilitate

[61] The new political parties emerging during the policy shift—FREPASO in Argentina, PRD in Mexico, and Causa R in Venezuela—built alliances with labor unions. Labor leaders, however, were more aware of the costs of corporatism once the state started its retrenchment, and the terms of new associations tended to be more fluid than in the past, thus affecting the extent of future "loyalties."

[62] Maria Victoria Murillo, "The Corporatist Paradox: Labor Parties and Labor Reform in Latin America" (Paper presented at the conference "Institutional Reforms, Growth, and Human Development in Latin America," Yale Center for International and Area Studies, April 16–17, 1999).

[63] Roberto Michels, *Political Parties: A Sociological Study of Oligarchical Tendencies in Modern Democracy*, trans. Eden and Cedar Paul (New York: The Free Press, 1962), 365.

[64] Ibid., 172.

[65] These implications are consistent with Hirschman's claim on the superiority of voice over exit as a mechanism for improvement in certain contexts when exit is not easily available or could provoke the demise of the organization.

the testing of alternative explanations within a relative small N. In turn a small N made possible the collection of the data necessary to tes causal mechanisms based on organizational dynamics. Theoreticall this research design demonstrates the possibility of holding national institutions and macroeconomic conditions constant in comparison within countries as well as sector-level variables constant in comparisons within sectors. Empirically it provides a better picture of the com plex reality of the countries studied and the organizational dynamics o each case while advancing comparative analysis.

APPENDIX:
CLASSIFICATION OF CASES

Venezuela			
Case	*Militancy*	*Concessions*	*Category*
CTV-1 1989–92 (after 1992 coup attempt market reforms are suspended)	First general strike in history in 1989; followed by other two general strikes, public demonstrations, and protests	Emergency wage hike, unemployment insurance, suspension of layoffs for six months, price controls for basic staples; more rigid labor law; no reform of severance payment system or social security institution; inclusion of workers' shares in privatization schemes	Opposition
Telecom. 1991 (privatization)	From five yearly strike petitions in 1989 and 1990 to only three in 1991	Employee-owned stock, labor directors, job stability, stability of collective bargaining contract	Cooperation
Telecom. 1991–93 (restructuring after growth of internal opposition)	In 1992 forty-two weeks were lost on wildcat and legal strikes in five unions of the national federation, and nine unions presented strike petitions; in 1993 twelve unions presented strike petitions	Restructuring only within the limits set by privatization agreement	Opposition

APPENDIX (*cont.*)

Venezuela			
Case	*Militancy*	*Concessions*	*Category*
Electricity 1989–91 (decentralization of the company)	No strikes in 1989, one in 1990, and none in 1991	Wage increases to the workers in the interior of the country, stability of work conditions across the company, handsome monetary incentives for transfers to the interior; job stability	Cooperation
Electricity 1992–93 (restructuring)	Four legal strikes and a wildcat strike in 1992; drastic increase in strike petitions, boycotts, and work unrest in 1993	Halt to restructuring attempts	Opposition
Oil 1991 (restructuring)	One strike in 1989, none in 1990, and one in 1991	Union allowed to retain hiring prerogatives in exchange for accepting the introduction of merit incentives on wages	Cooperation
Oil 1993 (restructuring)	Five strikes (one wildcat) in 1992 and six wildcat strikes in 1993	Union halts the dismantling of commissaries, but AD union leaders accept a reduction in hiring prerogatives demanded by management and internal competitors	Opposition
Auto 1989–92	No strikes or protests	Union accepts layoffs and introduction of new working methods in exchange for participation in selection of those laid off and trained in new technologies; after recovery, the union continues to participate in selection for training and reincorporation of laid-off workers	Cooperation

Venezuela			
Case	*Militancy*	*Concessions*	*Category*
Education	Strike petitions in the Ministry of Labor increased from nine in 1989 to thirty-one in 1990 and twenty-four in 1991 and dropped to four in 1992 and 1993; strikes also grew from three in 1989 to seven in 1990, eight in 1991, nine in 1992, and seven in 1993; occupation of buildings and street protests as well	No concessions administrative reform	Resistance

Mexico			
Case	*Militancy*	*Concessions*	*Category*
CTM-1	Fall in the number of strike petitions from a yearly average of 9,818 under the previous administration to a yearly average of 7,007	No concessions except halting the labor code reform	Subordination (except on the labor code where the unification of all PRI affiliates resulted in negotiation)
Electricity-1 1988–93 (attempt at privatization or liquidation of company)	From a 1987 strike (the first since 1936) under a previous opposition leadership to a public campaign, including a meeting of twenty thousand electrical workers, and lobbying to congress and to the president; public support for concertation pacts and NAFTA	Bail out of the company and creation of a new public company; monopoly of representation to the union; job stability; new fringe benefits and retirement plan; stability of work conditions and union participation in two union-management committees in charge of productivity and financial analysis; union right to	Cooperation

Mexico			
Case	*Militancy*	*Concessions*	*Category*
Electricity-2 1993–94 (restructuring)	Rejection of 1993 agreement signed by previous leader, first union to reject signing of the concertation pact in 1994; end of support for Salinas's policies; workers' mobilizations during bargaining in 1994; threats of strike	Reduction of the 1993 productivity targets in 1994, maintenance of job stability	Opposition (loyalty break)
Telecom.-1 1990–91 (privatization)	None	Employee-owned shares; labor director; increase in wages and permanent personnel; job stability; stability of work conditions	Cooperation
Telecom.-2 1992–94 (restructuring)	None	Committees for union participation in training and restructuring of work conditions, wage and benefits hikes	Cooperation
Oil (restructuring)	No strikes	No concessions and loss of job stability, work conditions, and union prerogatives	Subordination
Auto 1989–94 (restructuring)	Tradition of militancy in the 1970s, strike and sit-ins in 1989, plant takeover in 1990, none in 1991; demonstrations, sit-ins, meetings, stoppages, and "going naked" in 1993	Company-imposed conditions despite workers' resistance and sheer repression in 1990 and 1993	Resistance 1988–90 and 1992–93; Subordination 1990–91 and 1993–94

APPENDIX (*cont.*)

Mexico			
Case	*Militancy*	*Concessions*	*Category*
Education-SNTE (decentralization 1992)	Militancy by dissidents (CNTE) before 1989, but SNTE as well since 1989; militancy by SNTE peaked in 1991 against the fragmentation of the union and together with CNTE, including eight local strikes (besides four by the CNTE), three parades (besides three by CNTE), and many sit-ins, meetings, and a national process of consultation with the rank and file and joint public demands with the CNTE in November 1991	Limits to decentralization with guarantees of national work conditions and salaries by earmarking the state budgets, salary and benefits' hikes	Opposition

Argentina			
Case	*Militancy*	*Concessions*	*Category*
CGT 1 1989–92	From thirteen general strikes (1984–89) to one (1989–92); from 38.5 monthly strikes between January 1984 and June 1989 to 19.9 monthly strikes between July 1989 and March 1992	Unions could not stop hiring flexibility, wage restraint, and the loss of the collection of fees for their health funds	Subordination
CGT-2 1992–95	One general strike and 16.8 monthly strikes between April 1992 and July 1995	Union participation in the privatization of pensions, no reform to labor code, limits of competition for social security provision to health funds controlled by unions, permission for wage bargaining, union and workers' participation in privatization with government subsidies, bailing out of union debts	Cooperation

Argentina			
Case	*Militancy*	*Concessions*	*Category*
Electricity-1 1989–92 (privatization and restructuring)	Eleven yearly strikes in 1992 after a yearly average of less than three strikes in 1984–88	Employee-owned stock, labor directors, voluntary retirements, subsidies for union health fund; start-up contracts of privatized companies for union organized companies hiring laid-off workers	Opposition
Electricity-2 1992–95 (after expulsion of militant local unions)	No strikes by the national federation; (two yearly strikes in 1993–95 by expelled local union)	Subsidies for union participation in privatization of utilities and coal mines; union participation in company defining the spot price of electricity	Cooperation
Telecom.1 1989–90 (privatization and restructuring)	Nine yearly strikes in 1990 is the peak after an average of 4.5 yearly strikes during the previous administration	Employee-owned stock and labor director, voluntary retirements for layoffs, subsidy for the union health fund, executive appointments for the allied leaders	Opposition
Telecom.2 1991–94 (restructuring and outsourcing)	Drop in strikes to one yearly strike	Union participation in training, voluntary retirements for layoffs for the workers in the main companies	Cooperation
Oil-1 1990–91 (restructuring)	Drop in average yearly strikes from 4.6 in 1984–88 to 1.5 in 1990–91	Voluntary retirements for layoffs; subsidies to the union to organize its own health fund	Cooperation

Argentina			
Case	*Militancy*	*Concessions*	*Category*
Oil-2 1992 (privatization)	Yearly strikes averaged 0.5 in 1992–93	Union obtained monopoly of representation despite the existance of private oil workers' union; subsidies to the union to buy parts of the firm sold out by private owners with start-up contracts for hiring laid-off workers	Cooperation
Auto-1 1990–91 (between trade liberalization and the automobile regime)	Militancy increased from an average of 14.2 yearly strikes in 1984–89 to 18.5 yearly strikes in 1990–91	In 1991 unions and producers obtained a regime that protected the industry from competition	Opposition
Auto-2 1992–94 (automobile regime)	Militancy drops to 4.3 yearly strikes in 1992–94	Union and producers obtained the renewal of the automobile regime	Cooperation
Education-1 (decentralization law 1992)	Teachers' militancy was high with forty-four yearly strikes in 1989–94; proportion over total strikes increases to 44.5 percent in 1991 and 35.4 percent in 1992	No concessions	Resistance

SOURCES: Militancy data on Venezuela comes from the annual reports of the Secretary of Labor, a press chronology, and information of the labor relations' departments of CANTV, CADAFE, Ford Motors, and PDVSA. Militancy data on Mexico comes from a press chronology based on the archives of Entorno Laboral and from the Secretary of Labor. Strike data in Argentina is from the annual reports (1984–95) of the Consejo Técnico de Inversiones. Sources for concessions are labor contracts, government and union documents, press chronologies, and personal interviews with actors.

CONTINGENT DEMOCRATS
Industrialists, Labor, and Democratization in Late-Developing Countries

By EVA BELLIN*

DO social forces matter in democratic transition? Most of the recent literature on democratization would argue to the contrary. A surge of new research has focused on the role of elites and leadership, the importance of political institutions, and the consequences of strategic choice for democratic reform.[1] Democracy is variously portrayed in these studies as the crafted product of enlightened elites, the path-dependent yield of sticky institutions, or simply a conjunctural outcome. Although social forces are occasionally given the nod by inquiries into the strength of associational life[2] or the role played by coalitions,[3] this research has been overwhelmed by the flood of statecentric work that has dominated comparative politics for the past decade.[4] As Remmer points out, enthusiasm for bringing the state

* The author gratefully acknowledges the assistance of Frank Schwartz, Atul Kohli, Khaled Helmy, unihal Singh, Deborah Yashar, Susan Pharr, Sam Huntington, Tom Ertman, Dan Posner, and the mbers of the Sawyer Seminar/Research Workshop on Comparative Politics at Harvard University their comments on this research. Earlier versions of this paper were presented at the annual meet- , of the Middle East Studies Association Conference, November 1997, and the annual meeting of American Political Science Association, Atlanta, September 1999.

[1] See, for example, Guillermo O'Donnell and Philippe Schmitter, *Transitions from Authoritarian le: Tentative Conclusions about Uncertain Democracies* (Baltimore: Johns Hopkins University Press, 36); Guiseppe Di Palma, *To Craft Democracies: An Essay on Democratic Transitions* (Berkeley: Unisity of California Press, 1990); John Higley and Richard Gunther, *Elites and Democratic Consolida- 1 in Latin America and Southern Europe* (Cambridge: Cambridge University Press, 1992); Barbara ddes, "A Comparative Perspective on the Leninist Legacy in Eastern Europe," *Comparative Politi- Studies* 28 (July 1995); Arend Lijphart and Carlos H. Waisman, eds., *Institutional Design in New mocracies: Eastern Europe and Latin America* (Boulder, Colo.: Westview, 1996).

[2] See Robert Putnam, *Making Democracy Work* (Princeton: Princeton University Press, 1993); Au- ,tus Richard Norton, *Civil Society in the Middle East* (Leiden: E. J. Brill, 1995).

[3] See Deborah Yashar, *Demanding Democracy: Reform and Reaction in Costa Rica and Guatemala* anford, Calif.: Stanford University Press, 1997); Gregory Luebbert, *Liberalism, Fascism or Social mocracy: Social Classes and the Political Origins of Regimes in Interwar Europe* (New York: Oxford Unisity Press, 1991).

[4] Recent exceptions include Ruth Berins Collier, *The Contradictory Alliance: State, Labor Relations l Regime Change in Mexico*, University of California Research Series, no. 83 (Berkeley: University of lifornia, 1992); Ruth Berins Collier and James Mahoney, *Labor and Democratization; Comparing the st and Third Waves in Europe and Latin America*, Institute of Industrial Relations, Working Paper 62 (Berkeley: University of California, May 1995).

back in has pushed society out, largely excluding social forces from comparative analysis.[5]

It was not always so. A long tradition in political science put social forces—and more specifically *social classes*—center stage when explaining democratic outcomes. Both liberals and Marxists writing in the tradition of political economy long pointed to the protagonists of capitalist industrialization as the historical agents of democratization (although there was disagreement over which protagonist played the leading role). Focused primarily on the historical experience of the early industrializing countries of Western Europe, these scholars fell into two schools.

One school, led by Moraze, Hobsbawm, and Moore, identified the capitalist class as the class agent of democracy.[6] Moore summarized this view in his inimitable phrase "no bourgeoisie, no democracy." According to this school, West European democracy was the consequence of capitalists colliding with the absolutist state over the traditional, feudal barriers it posed to capitalist advance. Motivated by this material interest, capitalists mobilized their burgeoning economic power to create parliamentary institutions and impose parliamentary control over the state.[7]

A second school, led by Marshall, Thompson, Bendix, Therborn, and, most recently, Rueschemeyer, Stephens, and Stephens, identified the working class as the class agent of democracy.[8] According to this school, the political achievements of the capitalist class fell far short of democracy. Capitalists, they argued, were primarily interested in establishing liberal forms of rule, not democracy. While capitalists supported the introduction of representative government and the protection of civil liberties, they opposed the extension of political rights to the lower classes—which for these scholars is the mark of true democracy. In-

[5] Karen L. Remmer, "Theoretical Decay and Theoretical Development: The Resurgence of Institutional Analysis," *World Politics* 50 (October 1997), 57. See also Joel Migdal, Atul Kohli, and Vivienne Shue, eds., *State Power and Social Forces: Domination and Transformation in the Third World* (New York: Cambridge University Press, 1994).

[6] Charles Moraze, *The Triumph of the Middle Classes* (Garden City, N.Y.: Anchor Books, 1968); Eric Hobsbawm, *Industry and Empire* (Harmondsworth, England: Penguin, 1969); and Barrington Moore, *Social Orgins of Dictatorship and Democracy* (Boston: Beacon Press, 1966).

[7] James R. Kurth, "Industrial Change and Political Change: A European Perspective," in David Collier, ed., *The New Authoritarianism in Latin American* (Princeton: Princeton University Press, 1979).

[8] T. H. Marshall, *Citizenship and Social Class* (Cambridge: Cambridge University Press, 1950); E. P. Thompson, *The Making of the English Working Class* (New York: Pantheon, 1963); Reinhard Bendix, *Nation Building and Citizenship* (New York: Wiley Press, 1964); Goran Therborn, "The Rule of Capital and the Rise of Democracy," *New Left Review* 103 (1977); and Dietrich Rueschemeyer, Evelyne Huber Stephens, and John D. Stephens, *Capitalist Development and Democracy* (Chicago: University of Chicago Press, 1992).

stead, it was the working class, organizationally empowered by capitalist development and materially motivated by the desire to seek political redress for its economic subordination, that fought for universal suffrage. As such, the working class was the true champion of democratization in Western Europe.[9]

Despite their differences, however, both schools agreed on at least three central verities of democratic transition. (1) Democracy is neither an evolutionary necessity nor a conjunctural outcome; rather it is the product of *struggle* in which social forces play a central role. (2) *Interest*, not enlightenment, drives regime change. And (3) among the panoply of interests that animate people politically, *material* interests trump all others. This analysis suggests that social forces are most likely to champion democracy when their economic interests put them at odds with the authoritarian state.

But if capital and labor played an important role in championing democratic reform in the early industrializing countries of Western Europe (as the classic works of political economy argue),[10] the question is whether these social forces are likely to play a comparable role in the context of late development. A brief survey of late-developing countries shows wide variation in the enthusiasm of capital and labor for the democratic project. Organized labor has been an enthusiastic champion of democratization in some contexts (Korea, Chile, Zambia) but a more diffident partisan elsewhere (Mexico, Tunisia, Egypt). Private sector industrialists have championed democratization in some

[9] Revisionist historians have taken issue with any simple mythology that ascribes the rise of West European democracy to the work of a single, self-conscious social class, whether capitalists or workers; see Alex Callinicos, "Bourgeois Revolutions and Historical Materialism," *International Socialism* 43 (June 1989). Classes were internally divided, individual classes were often forced into coalitions with others to achieve political success, and other factors (institutional, international) also shaped regime change. The best works of political economy recognize the complexity of this process. In fact, a careful reading of Rueschemeyer, Stephens, and Stephens (fn. 8) reveals a nuanced argument along just these lines. See Thomas Ertman, "Democracy and Dictatorship in Interwar Europe Revisited," *World Politics* 50 (April 1998).

But acknowledging the importance of preexisting institutions and the agency of elites does not eny the central role played by social forces in democratic transition. Elites do not operate in a vacuum: heir political choices are governed not only by ideals and interests but also by social realities (for example, the economic pressure posed by the threat of capital flight; the political pressure posed by the hreat of organized and potentially destabilizing popular protest). And institutions are themselves a eation of the political process, subject to change in the face of political pressure and struggle. Deocratization itself is an example of institutional transformation, bought through the struggle of opsing interests that are equipped with different resources and agendas. The question is not so much *hether* social forces play a key role in the struggle for democratization as *which* social forces are likely take up the cause.

[10] Recent studies have challenged this position. See Collier and Mahoney (fn. 4), who argue perasively that labor played a much more negligible role in the first wave of democratic transition than t proposed by Therborn and others.

places (Korea by the mid-1980s, Brazil by the late 1980s) but have disdained it in others (Indonesia, Singapore, Syria). Can this variation be explained?

Level of economic development alone cannot account for it. Enthusiasm for democracy among industrialists and labor does not directly correlate with growth in per capita GNP, contra the expectations of old and new versions of modernization theory. In some cases private sector industrialists have proven most enamored of democratic transition precisely in moments of great economic decline (Brazil); in others, they have proven highly suspicious of democratization even in a context of spectacular economic growth (Indonesia, Malaysia). Similarly, some trade union movements have campaigned for democratization at times of great growth (Korea) while others have embraced it in a context of economic catastrophe (Zambia).

Cultural heritage also proves less than decisive. Comparisons drawn intertemporally within country cases as well as between countries with similar cultural endowments reveal dramatic variation in the enthusiasm of labor and industrialists for democracy, even as culture remains constant. Egypt and Tunisia, for example, possess similar cultural heritages (both are majoritarian Sunni Muslim countries, relatively unriven by ethnic cleavage, and saddled with a long history of West European colonialism), yet organized labor exhibits very different attitudes toward democratization in the two countries. Similarly, intertemporal comparison within the cases of Brazil and Korea shows a dramatic increase in industrialists' enthusiasm for democracy over the past fifteen years. Yet one would be hard pressed to argue that this shift was preceded by a revolution in the core cultural endowments of either country.

This article offers an alternative framework for explaining the variation in class support for democratization in the context of late development. Close study of a few core cases, in addition to more cursory examination of several others, helps specify the conditions under which capital and labor are more or less likely to embrace democratization. Although the article subscribes to the central verities of the political economy tradition, it argues that the peculiar conditions of late development often make capital and labor much more ambivalent about democratization than was the case for their counterparts among early industrializers. The theory will specify the conditions that give rise to this diffidence but will also suggest (and empirically anchor) the way these conditions may change to make both social forces more enthusiastic about democratic reform. The general lesson of this examination is

that capital and labor are *contingent* democrats[11] for the very reason that they are *consistent* defenders of their material interests. Like their predecessors, capital and labor in late-developing countries will champion democratic institutions when these institutions are perceived as advancing their material interests. But the pairing of material and democratic interest is contingent upon specific historical circumstances that are not necessarily replicated in the context of late development. Briefly put, capital and labor's enthusiasm for democracy largely turns on two variables. For capital, the two variables are state dependence and fear; for labor, they are state dependence and aristocratic position. Enthusiasm for democracy varies inversely with class score on these two variables. Where that score is split, class attitude toward democratization is indeterminate and other variables come into play.

To support this argument, the article begins with a discussion of the theoretical logic underlying the contingency hypothesis. It proceeds with empirical investigation of two core cases for capital (Indonesia and Korea) and two for labor (Mexico and Korea)—cases chosen for their dramatic variation on the dependent variable. Additional cases (Mexico and Saudi Arabia for capital; Egypt for labor) are added to correct for collinearity that appears in the two independent variables. Finally, a number of other cases (Tunisia, Brazil, and Zambia) are presented to eliminate rival hypothesis.[12]

Limited space prevents absolute parallelism in the cases investigated for capital and labor; it also precludes recounting the experience of capital and labor for every country mentioned (sixteen case studies in all, not counting intertemporal variation). Fortunately, the logic of comparison does not require exhaustive parallelism. Recounting the tale of some omitted cases (for example, Tunisian capital) would be largely redundant (its experience parallels that of Indonesian capital along our key variables) and would add little leverage to our hypothesis. Re-

[11] Guillermo O'Donnell describes the private sector's commitment to democracy as "contingent"; see O'Donnell, "Substantive or Procedural Consensus? Notes on the Latin American Bourgeoisie," in Douglas Chalmers, Maria de Souza, and Akko A. Boron, eds., *The Right and Democracy in Latin America* (New York: Praeger, 1992).

[12] To disprove the hypothesis that cultural heritage determines political disposition toward democracy, Tunisia and Brazil are added to the discussion of labor and capital, respectively. Using Mill's method of difference, comparison is drawn between cases showing overall similarity in cultural endowment (Tunisia and Egypt for labor; Brazil pre- and post-1980s for capital) but possessing a crucial difference in our independent variables, resulting in a crucial difference in outcome. To disprove the hypothesis that economic growth and prosperity determine political disposition toward democracy, Zambia and Brazil are also added to the discussion of labor and capital. Using Mill's method of agreement, comparison is drawn between cases showing overall difference in level of economic growth and prosperity (Zambia and Korea for labor; Brazil and Korea for capital) but sharing crucial similarity in their values on our independent variables, resulting in a striking similarity in outcome.

counting the tale of other omitted cases (for example, Saudi labor) would only confuse our discussion with exceptional conditions (for example, the fact that the vast majority of workers in Saudi Arabia are nonnationals with short-term horizons in the kingdom and little interest in shaping Saudi political institutions).

Careful case selection cannot eliminate all the methodological problems involved in drawing scientific inference here. The investigation of ten cases does not eliminate the problem of indeterminacy when four independent variables are proposed. Furthermore, case selection on the dependent variable introduces the possibility of biased results. As King, Keohane, and Verba observe, only *random* selection of *many multiple* cases can truly verify theory.[13] Nevertheless, the evidence presented seems sufficient to suggest the plausibility, if not the incontrovertibility, of the contingency hypothesis.

The Theory

State dependence, fear, and aristocratic position shape capital and labor's disposition toward democratization. Each of these variables merits elaboration. With regard to capital, state dependence refers to the degree to which private sector profitability is subject to the discretionary support of the state. Such support is typically delivered in the form of subsidized inputs, protected market position, close collaboration in the definition of economic policy, and state containment of labor and the capital poor. Two quite different state logics may fuel such support. Where the state is developmental (for example, Japan and Korea), it identifies national prosperity with that of the private sector and explicitly sponsors the development of the latter with the aim of achieving rapid economic growth.[14] By contrast, where the state is patrimonial (for example, Indonesia and Senegal), it also identifies prosperity with that of the private sector, but here the state's governing objective is not the achievement of rapid economic growth for the nation as a whole so much as personal gain for state elites.

But whatever the governing logic of the state, sponsorship makes the private sector diffident about democratization because capitalists recognize that their profitability hinges on state discretion. Therefore, in

[13] Gary King, Robert Keohane, and Sidney Verba, *Designing Social Inquiry: Scientific Inference in Qualitative Research* (Princeton: Princeton University Press, 1994), 115–49.

[14] The term "developmental state" was coined by Chalmers Johnson, *MITI and the Japanese Miracle* (Stanford, Calif.: Stanford University Press, 1982).

oth developmental and patrimonial cases, cozy collaboration with state lites, not public contestation and opposition, is the key to economic uccess. Where state elites are hostile to the project of democratic re- orm (and few authoritarian rulers relish the thought of relinquishing naccountable power), private sector entrepreneurs are careful to keep heir distance from it.

Private sector capital is also likely to be wary of democratization for easons that *do* vary with the character of the state. Where develop- nental logic rules, relations between the private sector and the state are ypically positive and collaborative. So long as the state generally antici- ates the interests of private sector capital, capital does not feel com- elled to create formal democratic institutions to ensure state ccountability. Where patrimonial logic rules, however, private sector rofitability typically resides in shady relations with state elites. Under hese conditions, transparency, one of the standard selling points of lemocracy, is likely to prove less attractive to entrepreneurs. In this way, collaborative profitability" whether governed by patrimonial or devel- pmental logic discourages enthusiasm for democratization among pri- ate sector capitalists.

Beyond the logic of collaborative profitability, the second variable hat hinders private sector enthusiasm for democracy is fear. Private ector capital everywhere is concerned, first and foremost, with protect- ng property rights and securing the long-term profitability of its in- estments through the guarantee of order.[15] But where poverty is videspread and the poor are potentially well mobilized (whether by ommunists in postwar Korea or by Islamists in contemporary Egypt), he mass inclusion and empowerment associated with democratization hreatens to undermine the basic interests of many capitalists. At best, uch inclusion threatens to flood politics with "the logic of distribution" ather than the "logic of accumulation."[16] At worst, it potentially con- ers upon the propertyless the means to overturn the social order. Thus, or many capitalists, democratization is associated with a deep sense of ocial threat and is regarded with distrust.

Why should state dependence and fear prove to be more significant arriers to democratic commitment among late-developing capitalists han among their predecessors?

[15] See Leigh Payne, *Brazilian Industrialists and Democratic Change* (Baltimore: Johns Hopkins Uni- rsity Press, 1994).

[16] See Thomas Callaghy, "Civil Society, Democracy and Economic Change in Africa: A Dissenting pinion about Resurgent Societies," in John Harbeson, Donald Rothchild, and Naomi Chazan, eds., *vil Society and the State in Africa* (Boulder, Colo.: Lynne Rienner, 1994), 243.

First, as Alexander Gerschenkron made clear,[17] late development heightens the dependence of capitalists on the state. Embarking on the process of industrialization in a world already industrialized, commercially integrated, and highly competitive means that private sector capitalists typically seek state help with capital accumulation (since the start-up capital for late industrializers often exceeds the capabilities of individual, first-generation entrepreneurs) and with trade protection (in the form of tariff barriers, import quotas, and so on). Such heightened state dependence makes private sector capitalists all the more wary of embracing political projects unpopular with state elites. By contrast, lower start-up costs and a less integrated international economy made the success of early industrializers significantly less contingent on state support, endowing capitalists with much greater political latitude.

Second, the developmental mission that guides state sponsorship in many late-developing countries also encourages private sector capitalists to cleave to the state and eschew the cause of democracy. Where the state is developmental, private sector capital can expect it to anticipate their interests to a large degree. This contrasts sharply with conditions faced by early industrializers whose feudal state was perceived to be hostile to capital. It was precisely this hostility that fueled capital's democratic conversion during the first transition. But in late-developing countries where the authoritarian state is often seen to be serving the interests of private sector capital, why embrace democracy?

Finally, one might wonder why capital would be more fearful of democracy today than during the first transition. Poverty after all was no less pervasive in eighteenth- or nineteenth-century England than it is today in many late-developing countries, and so the empowerment of the poor should have posed a comparable threat to capital's interests and made early capitalists equally leery of democracy. The difference, however, lies not in the relative number of poor people in society across time but rather in the degree of mass empowerment proposed by the democratic project in each era. During the first transition the hegemonic discourse on democracy had a distinctly liberal cast rather than an inclusionary one. Hence, capital's embrace of democracy was understood to be entirely consistent with exclusion of the propertyless. Since then, however, democracy has come to stand for mass inclusion, making it difficult for contemporary democratic discourse to justify exclusion based on property, race, or gender. The evolution in ideas, then,

[17] Gerschenkron, *Economic Backwardness in Historical Perspective* (Cambridge: Harvard University Press, 1962).

›es far toward explaining capital's divergent enthusiasm for democracy ross time. With democratization spelling mass empowerment today, any capitalists in late-developing countries are more diffident about nbracing this political project.

What about organized labor? Labor's enthusiasm for democracy rns largely on two variables: state dependence and aristocratic posi›n. State dependence refers to the degree to which organized labor ·pends on state support for its organizational viability, vitality, and ›ut. Labor everywhere, with its reliance on *collective* action to estabh influence, is much more dependent upon the state than is private :tor capital.[18] But beyond this run-of-the-mill dependence, labor's rence on the state can be deepened by the state's adoption of a corpotist strategy that provides unions with financial and organizational pport in exchange for political loyalty and self-restraint. This rangement provides labor with material benefits far in excess of what true market power can deliver—but at the price of its autonomy.

Labor's dependence upon the state gives rise to diffidence about deocratization for reasons similar to those that operate in the case of pital. Labor comes to fear biting the hand that feeds it, that is, jeopdizing the flow of state benefits by embracing political projects that : certain to evoke the wrath of state elites. To the contrary, organized, ıte-dependent labor believes its interests are better served by mainning collaborative, not contestatory, relations with the state.

The second variable modulating labor's enthusiasm for democracy is stocratic position—the degree to which organized labor is economilly privileged vis-à-vis the general population. Where organized labor joys such a privileged stance, it is likely to exhibit a degree of "dislidarity" with the unorganized masses in the informal sector and/or riculture. Under such conditions, and especially where labor's aristotic position is a consequence of political intervention rather than a lection of true market power, labor will perceive its interests to be tter served by maintaining a cozy relationship with the state (even if : institutional arrangements are authoritarian), rather than by cham›ning institutions that make the state accountable to mass interests ıat is, democracy).

Again, conditions of late development heighten the probability that ›or will be state dependent and/or aristocratic and hence more diffint about the democratic project. Late development typically spells uctural weakness for labor because it is accompanied by industrial-

' Claus Offe and Helmut Wiesenthal, "Two Logics of Collective Action," in *Political Power and al Theory* (Greenwich, Conn.: JAI Press, 1980).

ization in a context of global capital mobility. This situation universalizes the "reserve army" of labor and undermines the bargaining power of workers. Late timing also means the importation of industrial technology that tends to be capital intensive rather than labor intensive, and this exacerbates the problem of labor surplus and further weakens labor's market position. Such weakness encourages labor to look to the state for political remedies, fostering union dependence and nurturing diffidence about contestatory politics. At the same time, the problems of labor surplus and mass poverty mean that the minority of workers who are organized in the formal sector are likely to be privileged vis-à-vis the majority, an economic privilege jealously guarded by organized labor. The result is that the interests of the unions are quite distinct from those of the unorganized poor, and this discourages the unions from fighting for state accountability to mass preferences.

Of course none of this is cast in stone. Economic growth may absorb labor surplus and diminish labor's structural weakness and dependence upon the state. Economic growth and/or welfarist public policy may reduce mass poverty and diminish capital's sense of social threat. Fiscal crisis or political exigencies internal to the authoritarian state may lead it to reduce its sponsorship of private sector capital, leading capital to reconsider the advantage of coziness over formalized accountability as the surest route to profitability. Under such conditions capital and labor may reconsider the advantages offered by democratic reform. For labor, democracy holds out the promise of civil liberties such as freedom of speech and association—the bedrock of collective action and collective power. For capital, democracy holds out the promise of institutionalized accountability and transparency—the means to more predictable influence over policy for the well organized and the well heeled. Thus, after weighing the costs and benefits, capital and labor may be converted to the democratic cause. But their commitment is always refracted through the prism of interest and can be predicted only on the basis of a clear understanding of this interest and the variables that shape it.

The power of our variables to explain class commitment to democracy will be explored in depth in two core cases each for capital and labor, with more cursory corroboration drawn from six other country cases. Capital's dependence on the state will be measured in terms of dependence on subsidized inputs, protected markets, and cronyistic relations with state elites. Capital's sense of fear or threat will be measured in terms of the pervasiveness of poverty, the organizational

strength of the capital poor (for example, membership levels in the Communist Party), and past incidence of popular violence. Labor's dependence on the state will be measured in terms of union dependence on state subsidies and union members' access to state-subsidized benefits (for example, credit and housing), as well as the politically manipulated (inflated) setting of wage levels. Labor's aristocratic position will be measured in terms of differentials found between the organized and unorganized in matters of wage levels, access to stable employment, social security, and other nonwage benefits such as legally mandated job security. While these variables are not exhaustive, they anticipate a great deal of the variation found in class commitment to democracy, both among cases and even more powerfully *within* cases across time. A schematic summary of the argument and positioning of the cases is presented in Figures 1 and 2.

		State Dependence	
		High	Low
Fear	High	**Antidemocracy** Indonesia	**Leaning Democratic** Mexico (by the 1970s)
	Low	**Ambivalent** Saudi Arabia	**Prodemocracy** Korea (by mid-1980s) Brazil (by the 1980s)

FIGURE 1
PRIVATE SECTOR CAPITAL: CHAMPION OF DEMOCRACY?

		State Dependence	
		High	Low
Aristocratic Position	High	**Antidemocracy** Mexico (CTM) Tunisia	**Leaning Democratic** Null
	Low	**Ambivalent** Egypt	**Prodemocracy** Korea Zambia

FIGURE 2
ORGANIZED LABOR: CHAMPION OF DEMOCRACY?

CAPITAL AND DEMOCRATIZATION

INDONESIA: PATRIMONIAL CAPITALISM, FEAR, AND DEMOCRATICALLY DIFFIDENT CAPITALISTS[19]

Until relatively recently Indonesia was celebrated for its success as one of Southeast Asia's most rapidly growing "tigers." But like many other successful Asian countries, Indonesia put the lie to any assumption that economic growth necessarily spells enthusiasm for democracy among the social classes that are its primary beneficiaries.[20] Specifically, private sector capitalists in Indonesia have proved consistently reluctant to embrace projects to democratize the country. To explain the business community's alliance with authoritarianism, we must explore the historic character of the state's relationship with the private sector and the prevalence of social fear within the business community.

From independence onward the Indonesian state consistently championed the development of private sector capital, especially the development of an *indigenous* (that is, Malay) capitalist class. To this end, the state gave indigenous entrepreneurs preferential access to essential business benefits such as government contracts, licenses, bank credit, and trade protection. It put into effect an "entrepreneurial affirmative action program" designed to boost the number of indigenous entrepreneurs (*pribumi*) by legislative fiat. And it embraced the practice of "bureaucratic capitalism," which permitted public officials to become entrepreneurs even while in office.[21]

State sponsorship of private sector development was patrimonial in nature. Public officials distributed government support to private firms with an eye to securing personal profit and a political clientele. Officials doubling as entrepreneurs used their control over the allocation of licenses, concessions, and credit to promote their own companies—blurring the boundary between the public and private sectors. Nevertheless, the private sector flourished and grew dramatically over the first four decades of Indonesian independence.

[19] This analysis draws extensively on Richard Robison, *Power and Economy in Suharto's Indonesia* (Manilla, Philippines: Journal of Contemporary Asia Publishers, 1990); Kevin Hewison et al., eds., *Southeast Asia in the 1990s: Authoritarianism, Democracy, and Capitalism* (St. Leonards, Australia: Allen and Unwin, 1993); Richard Robison, *Indonesia: The Rise of Capital* (Sydney: Allen and Unwin, 1986); and Andrew MacIntyre, *Business and Politics in Indonesia* (Sydney: Allen and Unwin, 1991).

[20] See Jacques Bertrand, "Growth and Democracy in Southeast Asia," *Comparative Politics* 30 (April 1998); David Martin Jones, "Democratization, Civil Society, and Illiberal Middle Class Culture in Pacific Asia," *Comparative Politics* 30 (January 1998).

[21] Robison (fn. 19, 1986), 167; Hewison et al. (fn. 19), 46.

The character of state sponsorship explains much of the private sector's diffidence about democratization. The fact that licenses, contracts, and credit were distributed on a discretionary basis, with access governed by political logic (or official gain), rather than by publicly formulated, economically rational criteria, meant that entrepreneurs were wise to nurture cozy relations with state elites. The fact that much of the business collaboration between state elites and private sector actors was quite shady in nature, if not outrightly corrupt, made political transparency (a good associated with democracy) less attractive. The fact that many officials doubled as entrepreneurs reassured the business community that state elites would anticipate private sector interests when formulating public policy, obviating the need for more formal mechanisms of accountability. Private sector capitalists thus had reason to be diffident about championing democratization. So long as the state continued to deliver economic prosperity, private sector capital had little incentive to push for political reform.

But if dependence on state support fostered private sector diffidence about democratization, then so did fear, which for the business community in Indonesia had three underlying components. First, there was the problem of widespread poverty and the fact that mass empowerment had historically been associated with threats to the social order. During the late 1950s and early 1960s more than 60 percent of Indonesians lived below the poverty line.[22] Pervasive poverty helped fuel the popularity of a strong Communist Party, with a membership approaching three million by the mid-1960s;[23] by 1965 the party had been implicated in a regime-threatening coup. The sense of revolutionary danger that pervaded Indonesia at the time turned the propertied classes against democratic experiments and pushed them into the arms of authoritarian stability. This was a historic legacy that lingered.

But even after the Communist Party had been decimated by the military in 1965 and rapid economic growth had reduced mass poverty in the 1970s and 1980s, the business community still had reason to fear mass empowerment. From the early 1980s onward the regime's development strategy turned on the exploitation of a docile labor force—which required containment of the masses, not their empowerment.

[22] Hal Hill, ed., *Indonesia's New Order* (Honolulu: University of Hawaii Press, 1994), 57; World Bank, *World Bank Development Report* (New York: Oxford University Press, 1990), 41–43.

[23] The party also organized upwards of sixteen million citizens in its mass organizations of peasants, trade unionists, women, and youth, and in the 1955 national elections it won 16.4 percent of the popular vote, making it one of the big four parties in Indonesia. Brian May, *The Indonesian Tragedy* (Boston: Routledge, 1978), 113.

The growing labor unrest of the 1980s and 1990s only reinforced capital's belief that a repressive state apparatus was essential to economic success.[24]

Finally, the business community's fear of mass empowerment had a third component, namely, ethnic conflict and the tendency for class warfare to take on ethnic coloration in Indonesia. Because the vast majority of domestic capitalists hail from the Chinese minority (while most Indonesians are Muslim Malay), class discontent over inequity and exploitation is frequently channeled along ethnic/religious lines, with protest expressed in the language of Islamic populism and riots directed against the Chinese community as a whole.[25] Violence against the Chinese community has been a recurrent problem, growing in intensity since the 1950s and culminating most recently in the May 1998 "orgy of looting, plundering, and fire-raising" aimed at Chinese businesses.[26] As a result, the Chinese suffer from a sense of social vulnerability that has long led many Chinese capitalists to prize stability over freedom and a strong authoritarian state over popular empowerment.

For these reasons then, private sector capital in Indonesia has historically proven unenthusiastic about democratization. The private sector's dependence upon the state for its profitability, as well as its fear of mass empowerment have long allied it with the authoritarian order.[27] Nor have recent events prodded the political conversion of the private sector. Catastrophic economic crisis in 1997–98 fueled regime change in Indonesia and a tentative transition toward democracy. But it also quadrupled mass poverty (eighty million now live below the poverty line) and sparked violent interthnic strife.[28] Both have fanned the business community's traditional fears about mass empowerment and this, together with concerns about the new regime's (IMF-supported) attacks on business-state cronyism, have left business leaders skeptical about, if not actually hostile to, Indonesia's fragile new regime.[29]

[24] R. William Liddle and Rizal Mallarageng, "Indonesia in 1996," *Asian Survey* 37 (February 1997).

[25] Hewison et al. (fn. 19), 58–60; Robison (fn. 19, 1986), 274–76, 315–20; MacIntyre (fn. 19), 3.

[26] *Country Report, Indonesia, Economist* Intelligence Unit (2d quarter, 1998), 19. For accounts of earlier anti-Chinese violence, see "Anti-Chinese Outbreaks in Indonesia, 1959–68," in J. A. C. Mackie, *The Chinese in Indonesia* (Sydney: Australia Institute of International Affairs, 1976).

[27] Robison (fn. 9, 1986), 150–52; Jacques Bertrand, "Business as Usual in Suharto's Indonesia," *Asian Survey* 37 (May 1997), 443.

[28] Adam Schwartz and Jonathan Paris, eds., *The Politics of Post-Suharto Indonesia* (New York: Council on Foreign Relations, 1999), 2–9.

[29] *Country Report, Indonesia, Economist* Intelligence Unit (3d quarter, 1998), 13.

SOUTH KOREA: VARIATION PROVES THE RULE[30]

The South Korean case, like that of Indonesia, confirms the importance of state dependence and fear for shaping the political disposition of private sector capital in the context of late development. But ultimately, the Korean case provides more compelling evidence of this relationship thanks to its internal variation. During the first few decades of the postwar era, the high level of state dependence and the prevalence of fear within the Korean business community spelled private sector diffidence toward democratization. From the mid-1980s, however, political and economic developments in the international and domestic arenas signaled important changes in the character of state sponsorship, the extent of state dependence, and the prevalence of social fear in the business community. Coincident with these changes, private sector capitalists began to exhibit significant enthusiasm for political reform and democratization. This variation over time provides compelling confirmation of the role played by our two variables in shaping the political inclinations of private sector capital. It also testifies to the intrinsically contingent nature of the political leanings of private sector capital.

For the sake of conciseness and because the logic of this period largely replicates that of the Indonesian case, the analysis of South Korean capital's first four postwar decades is given in brief. At this time the character of state sponsorship played a key role in fostering private sector diffidence about democratization. During the patrimonial rule of Syngman Rhee both the genesis and the success of private sector firms turned on political mediation and access to the state's discretionary favor. Cronyism governed the start of many private sector firms.[31] And preferential access to government-controlled resources (including tax breaks, trade monopolies, foreign-aid disbursement, and low-interest loans) governed private sector growth.[32] Neither nurtured a culture of political contestation or kindled a desire for transparency among

[30] This analysis draws extensively on Carter Eckert, "The South Korean Bourgeoisie: A Class in Search of Hegemony," *Journal of Korean Studies* 7 (Fall 1990); idem, *Korea Old and New* (Cambridge: Ilchokak, Korea Institute, Harvard University, 1990); Hagen Koo, *State and Society in Contemporary Korea* (Ithaca, N.Y.: Cornell University Press, 1993); Karl Fields, *Enterprise and the State in Korea and Taiwan* (Ithaca, N.Y.: Cornell University Press, 1995); Nora Hamilton and Eun Mee Kim, "Economic and Political Liberalisation in South Korea and Mexico," *Third World Quarterly* 14, no. 1 (1993); George Ogle, *South Korea: Dissent within the Economic Miracle* (London: Zed Press, 1990).

[31] Eckert (fn. 30, 1990), 40; Ogle (fn. 30), 42.

[32] Fields (fn. 30), 31–40; Minho Kuk, "The Governmental Role in the Making of Chaebol," *Asian Perspective* 12, no. 1 (1988), 110–17.

Korea's business elites. Later, under Park Chung Hee the state adopted a more developmental ethos, but this did not reduce the private sector's dependence. The state retained discretionary control over the distribution of resources, such as credit and foreign currency, that were critical to the private sector, and this continued to nurture a culture of compliance within the business community, as well as a lack of interest in political campaigns that might jeopardize the flow of state benefits.[33]

The state's conversion to a developmental ethos fostered private sector diffidence about democratization in a second way as well. By prioritizing economic development and identifying public welfare with private sector growth, the state proved extremely attentive to capital's interests. Business leaders thus felt little need to push for more formal mechanisms of accountability.

But the private sector's lack of interest in democracy at this time was also fueled by fear, specifically fear of the revolutionary potential of mass empowerment. Korea emerged from the Korean War an impoverished country (an estimated ten million were without homes, adequate food, or medical care),[34] and its poor had historically found revolutionary inspiration (and organizational wherewithal) in a popular and activist Communist Party. Although actual membership in the Korean Communist Party remained small (at between forty and sixty thousand members), the party's extraordinary capacity to organize citizens in a network of youth, labor, and peasant organizations meant it could mobilize hundreds of thousands, if not millions, for communist purposes.[35] Enfranchising the poor thus had the potential to threaten the capitalist social order. The victory of communism in North Korea and the continuing state of war between North and South only heightened the sense of danger felt by capital interests. Private sector capitalists were also disinclined to embrace mass empowerment because of its potential to undermine Korea's development strategy. Like Indonesia in the 1980s and 1990s, Korea in the 1950s and 1960s embraced a strategy of export-oriented industrialization whose success turned on the repression and exclusion of labor. Democratization, which potentially might empower labor, loomed as an unwelcome prospect for Korean capitalists.

[33] State monopoly of the banking system (1961–80) and state supervision of access to foreign loans and grants endowed it with substantial financial power; Fields (fn. 30), 95–96, 121.

[34] Parvez Hasan, *Korea* (Baltimore: Johns Hopkins University Press, 1976), 26. In 1953 Korea was one of the poorest countries in the world with a GNP per capita of $67 (in 1996 dollars); Byong-Nak Song, *The Rise of the Korean Economy* (Hong Kong and New York: Oxford University Press, 1994), 60.

[35] Robert Scalapino and Chong-sik Lee, *Communism in Korea*, vol. 1 (Berkeley: University of California Press, 1972), 268.

For the better part of the postwar era, then, Korean capitalists closed ranks behind the authoritarian state. By the mid 1980s, however, the political attitudes of capital had shifted, moving away from a culture of passivity and consent and toward frank embrace of political contestation and democratization. The first glimmers of this shift came in 1986–87, when the leading business confederation (the Federation of Korean Industries) advanced a new political agenda, boldly independent of the ruling party.[36] More dramatic evidence came in 1991, when Chung Ju Yung, the founder and chair of the Hyundai group, one of Korea's largest chaebol, established a new political party designed to challenge the ruling DJP. The party managed to win 17 percent of the popular vote in the general elections of 1992, and later that year Chung even made a bid for the presidency.[37] Such assertiveness marked an important departure for capital.

Much of this new political orientation away from complicity in authoritarian rule and toward embrace of democratic contestation and pluralism is explained by the reduction in the private sector's dependence upon, and receipt of, state support, in addition to a general decline in social fear. During the 1980s and 1990s two factors led the Korean state to reduce its sponsorship of private sector industry. First, the worldwide recession engendered by the oil crisis of 1979 created serious difficulties for Korea's highly indebted, trade-dependent economy. The Chun regime was forced to adopt a wide-ranging program of structural reform that decreased state support for the private sector. Credit supplies were reduced, policy loans were eliminated, protection of the domestic market was decreased, and a new trade law ended the long-standing monopoly positions enjoyed by many chaebol in the domestic market.[38] The regime's commitment to economic reform eliminated many of the rents that had long sustained domestic capital in Korea and marked a novel disjuncture in the interests of state and capital.

Second, the country's shift to competitive electoral politics in 1987 created a new political imperative for the regime to woo popular support through political and social reform.[39] As part of its campaign to

[36] Eckert (fn. 30, 1990).

[37] Koo (fn. 30), 47; Fields (fn. 30), 60.

[38] Stephan Haggard and Chung-in Moon, "The State, Politics, and Economic Development in Postwar South Korea," in Koo (fn. 30).

[39] Eckert (fn. 30, 1990), 377–79. The regime's decision to introduce competitive elections came in response to widespread popular demonstrations for democracy in 1986–87. The pending summer Olympic games, scheduled to be held in Korea in 1988, also subjected the regime to greater international scrutiny and heightened the effectiveness of popular protest. The business community was not at the vanguard of this popular movement. See Hamilton and Kim (fn. 30), 119–20.

build a popular base, the regime sought to distance itself from big business by attacking the chaebol for corruption and reining in many of their long-standing privileges. The state imposed stricter controls on bank credits and raised inheritance taxes, policies that infuriated the business commuity.[40] In short, structural adjustment and political opening meant that the state would no longer be so reflexively solicitous of business interests. To the contrary, the state directly trampled on business interests, leading many domestic capitalists to view it with increasing distrust. This encouraged business leaders like Chung of Hyundai to look for new ways to make the state more responsive to their interests.

Alongside the reality of declining state support, the private sector's declining *need* for that support also made it more receptive to democratization. During the first postwar decades Korea's late industrialization had spelled private sector dependence upon the state for capital accumulation, entrepreneurial direction, and protection of the local market. But by the mid-1980s many of the grounds for private sector dependence had been eliminated. Korean industry was competitive enough so that state protection from foreign rivals was no longer essential. Korean capital had developed a deep store of entrepreneurial experience, rendering state direction increasingly superfluous.[41] And many of the chaebol were sufficiently large to provide on an in-house basis many of the financial services that the state had previousl supplied.[42] This, together with the liberalization of the banking sector in 1980, the creation of nonstate financial institutions, and the internationalization of financial markets, meant that the business community had access to independent sources of corporate financing beyond the states's control.[43] By the late 1980s the private sector had overcome many of the dependencies associated with latecomer status, and this decreased dependency created new opportunities for capital to break ranks with the regime and embrace political reform.

Finally, a general decline in fear of mass empowerment made business more sympathetic to democratization. By the late 1980s expanding prosperity had decreased the sense that capital and labor were locked in a zero-sum conflict.[44] Although income inequality remained signifi-

[40] Koo (fn. 30), 48.

[41] See Hamilton and Kim (fn. 30), 116; Koo (fn. 30), 47.

[42] Hamilton and Kim (fn. 30), 118.

[43] Koo (fn. 30), 88.

[44] Tat Yan Kong, "Origins of Economic Liberalization and Democratization in South Korea," in Gerd Nonneman, ed., *Political and Economic Liberalization* (Boulder, Colo.: Lynne Rienner, 1996), 241.

cant,[45] absolute poverty had largely been eliminated, with less than 10 percent of the population living below the poverty line.[46] Demands for material improvement no longer aroused fears of upheaval. In addition, the global decline of communism reduced the revolutionary implications of popular empowerment, such that domestic mass movements with social welfare agendas were no longer perceived as the foot soldiers of an international revolutionary project. Consequently, capital could be persuaded that mass exclusion was no longer essential to preserving the social order.

Nevertheless, one should not exaggerate the extent of capital's conversion to democracy. Its commitment remains contingent on democracy's consistency with its economic interests. And some of these interests (for example, fear of higher wages and lowered profitability of exports) still incline capital toward political repression, especially the containment of labor. But thus far, at least, an overall calculation of interest has not induced capital to abandon support for democracy.

Corroboration from Brazil, Mexico, and Saudi Arabia

The conclusions suggested by the Korean and Indonesian cases for the logic governing capital's contingent commitment to democracy are corroborated by other cases. The Brazilian case, for example, replicates both the intertemporal variation found in Korea, as well as its etiology.[47] Brazil's private sector was closely allied with the country's embrace of authoritarianism in 1964, but by the late 1980s much of the private sector had come to endorse democratization. The reasoning behind this change of heart echoes that of the Korean case.

During the 1960s the private sector was gripped by fear of revolutionary insurgence. Widespread poverty, a radical left eager to mobilize the economically disadvantaged, and a cold war context gave substance to the threat of insurgence. In addition, the private sector, highly dependent on the state for subsidies, contracts, credit, and technology, sought to empower a state that would be responsive to its interests and committed to its growth. Both factors motivated capital to support an

[45] Hagen Koo, "The Political Economy of Income Distribution in South Korea," *World Development* 12, no. 10 (1984), 1030–31. Already by 1978 absolute poverty had declined to 12 percent, down from 41 percent in 1965.

[46] Andrea Savada and William Shaw, eds., *South Korea: A Country Study* (Washington, D.C.: Library of Congress, 1992), 177.

[47] This analysis draws extensively on Payne (fn. 15); Fernando Cardoso, "Entrepreneurs and the Transition Process: The Brazilian Case," in Guillermo O'Donnell et al., *Transitions from Authoritarianism* (Baltimore: Johns Hopkins University Press, 1986); Leigh Payne and Ernest Bartell, eds., *Business and Democracy in Latin America* (Pittsburgh, Pa.: University of Pittsburgh Press, 1995); and O'Donnell (fn. 11)

authoritarian regime that promised containment of the left and sponsorship of the private sector.[48] By the late 1980s, however, changing conditions had led the private sector to reevaluate its support for authoritarianism. Fear of revolutionary insurgence had dissipated in light of the change in international conditions (the demise of the Soviet Union, the end of the cold war) and the deradicalization of the left after two decades of repression. In addition, the authoritarian regime had proven less than fully responsive to capital's interests, privileging public enterprises over private and expanding the state's regulation of the economy at the expense of private sector interests.[49] Thus, both declining fear and declining state support persuaded the private sector to abandon the authoritarian regime and support democratic transition. In this way Brazil's private sector constitutes an archetypical case of contingent democrats. Like their Korean counterparts, they shifted toward an embrace of democracy on the basis of a "utilitarian calculus" of their material interests.[50]

Note that the Brazilian case also helps eliminate rival hypotheses to our theory of democratic contingency. As in Korea, intertemporal comparison within the Brazilian case shows significant variation in private sector enthusiasm for democracy, despite constancy in the country's core cultural endowment. This suggests that cultural heritage alone cannot account for variation in class enthusiasm for democracy. Further, comparison of the Brazilian case with the Korean case reveals significant similarity in private sector enthusiasm for democracy, despite variation in economic context. (Brazil faced economic crisis at the moment of private sector enthusiasm for democracy, whereas Korea was experiencing rapid economic growth.) This suggests that level of economic prosperity alone cannot anticipate variation in class enthusiasm for democracy.

All the cases presented thus far manifest a degree of collinearity between our two independent variables. It is possible, however, to find cases where the two variables do not covary. Where this is true, preliminary evidence suggests that, for capital, state dependence may be a more powerful inhibitor of enthusiasm for democracy. The cases of Mexico and Saudi Arabia support this impression.

Mexico provides a case of *declining* state dependence but *persistently high* social fear for capital. Under these conditions private sector capital

[48] The Goulart regime that existed prior to the 1964 coup was perceived to be unpredictable, incompetent, and inattentive to private sector interests. See Payne (fn. 15), 13

[49] Cardoso (fn. 47), 143.

[50] O'Donnell (fn. 11).

has been persuaded to embrace democratization.[51] The private sector was long the dependent stepchild of the Mexican revolution. Excluded from any public role in politics but generously nurtured by the state's "alliance for profits," the private sector long acquiesced in the regime's authoritarianism. Beyond dependence on state support, the private sector's alliance with authoritarianism was further fueled by fear. Although the threat of a radical left had largely been shut out by the ruling revolutionary party, the mass poverty and income inequality found in Mexican society were grounds for concern among the propertied classes and reason to support an exclusionary, even repressive, state apparatus.

By the late 1970s and early 1980s, however, the rationale for capital's alliance with the authoritarian regime began to unravel. Not that the sources of private sector fear—massive poverty or income inequality—had decreased. If anything these scourges worsened as currency problems and fiscal deficits plunged Mexico into severe economic crisis during the 1980s.[52] Rather, it was the private sector's economic dependence on the state (and its confidence in state sponsorship) that began to decline. At this time there emerged a new group of private sector entrepreneurs who were export oriented and less dependent on state support and protected markets for their prosperity. They resented the state's corrupt intervention in the economy and were vexed by many of its policy decisions, notably the nationalization of banks during the currency crisis of 1982. As private sector entrepreneurs began to question the state's "alliance for profits," an important segment of them also "discovered democracy."[53] They began to push for the democratization of Mexico's political system and spearheaded political pluralization by bankrolling the political party PAN. In the Mexican case decreased state dependence but invariable social threat spelled private sector support for democratization.

In Saudi Arabia, by contrast, state dependence has remained *persistently high* but social threat has *declined.* In this context the private sector has remained largely ambivalent about democratization. The dependence of the private sector on the Saudi state is legendary. Indus-

[51] This analysis draws extensively on Susan Kaufman Purcell, "Business-Government Relations in Mexico: The Case of the Sugar Industry," *Comparative Politics* 13 (January 1981); Roderic Camp, *Entrepreneurs and Politics in Twentieth-Century Mexico* (Oxford: Oxford University Press, 1989); Sylvia Maxfield and Ricard Anzaldua, eds., *Government and Private Sector in Contemporary Mexico* (San Diego: Center for U.S.-Mexican Studies, University of California at San Diego, 1987).

[52] See Sebastian Edwards, *Crisis and Reform in Latin America: From Despair to Hope* (New York: Oxford University Press, 1995); Kevin Middlebrook, ed., *Union, Workers, and the State in Mexico* (San Diego: University of California at San Diego Press, 1991).

[53] The phrase is borrowed from Cardoso (fn. 47).

trial ventures are highly protected and subsidized by the government. Many manufacturing firms register *negative* value added but survive nonetheless thanks to state guarantees of preset profit margins for private sector ventures.[54] The state's "allocative logic"[55] sustains these firms just as it has eliminated a major source of social fear in the Saudi kingdom. Since the mid-1970s the state has funded massive social welfare projects in the country, investing billions of dollars in health, education, housing, and social safety nets to eliminate the problem of mass poverty. Despite this declining social threat, however, the private sector has not judged it advantageous to champion democratization.[56] Although leaders of the business community have expressed an interest in strengthening the rule of law in the kingdom, their political ambitions have fallen far short of democratization. Thus, even as businessmen have circulated petitions calling for due process and a strengthened judicial system, they have continued to declare their fidelity to the Saudi royal family, as well as to the monarchical system of government. Clearly, business elites do not wish to jeopardize their access to state largesse, and this has muted any impulse to push for extensive democratic reform.[57]

As the cases presented thus far demonstrate, private sector capital has shown wide variation in its support for democracy, both across cases and within cases across time. Two variables—degree of state dependence and level of social fear—explain a good deal of this variation. Overall, the empirical evidence suggests an inverse relationship between variables and outcome. That is, as the levels of state dependence and social fear decline, the likelihood that the private sector will embrace democracy increases. Where the score on these two variables is

[54] See Hazem Beblawi and Giacomo Luciani, *The Rentier State* (New York: Croom Helm, 1987). The dependence of private sector entrepreneurs in nonindustrial ventures (e.g., commerce, real estate, construction) is even more legendary. For a colorful account, see Michael Field, *The Merchants: The Big Business Families of Saudi Arabia and the Gulf States* (New York: Overlook Press, 1984).

[55] Giacomo Luciani, "Allocation vs. Production States: A Theoretical Framework," in Luciani, ed., *The Arab State* (Berkeley: University of California Press, 1991).

[56] See Helen Chapin Metz, ed., *Saudi Arabia: A Country Study* (Washington D.C.: U.S. Department of State, 1993). Of course, the elimination of mass poverty has not eliminated all sources of social threat in Saudi Arabia. Expatriate workers (largely excluded from the state's social welfare benefits), the Shiite minority (which suffers second-class citizenship), and Islamist activists all constitute potential sources of insurrection. Neither of the latter two groups, however, constitutes a popular majority who might use mass empowerment to overturn the status quo. Expatriate workers are not citizens and so would not be empowered by democratization. See R. Hrair Dekmejian, "The Rise of Political Islamism in Saudi Arabia," *Middle East Journal* 48 (Autumn 1994); Mahdawi al-Rasheed, "Saudi Arabia's Islamic Opposition," *Current History* 95 (January 1996), 16–22; Peter Wilson and Douglas Graham, *Saudi Arabia: The Coming Storm* (New York: M. E. Sharpe, 1994).

[57] On Saudi elites' limited demands for political reform, see F. Greg Gause III, *Oil Monarchies: Domestic and Security Challenges in the Arab Gulf States* (New York: Council on Foreign Relations, 1994).

split, state dependence proves the more powerful inhibitor of capital's enthusiasm for democracy.

Organized Labor and Democratization

If the preceding empirical investigation makes clear the contingent and instrumental nature of capital's commitment to democratization, further empirical study confirms a similar contingency and instrumentality on the part of organized labor. The two variables governing labor's contingent support for democracy are state dependence and aristocratic position. Again, the evidence suggests an inverse relationship between these variables and enthusiasm for democracy.

Mexico: State Dependence, Aristocratic Privilege, and Democratically Diffident Labor[58]

The Mexican case provides compelling evidence of the *diffident* attitude organized labor may exhibit toward democratization. The leading component of organized labor, the CTM, remained steadfastly allied with the Mexican regime even as that regime's restrictions on civil liberties, repression of the opposition, and repeated noncompetitive elections attested to its authoritarian character. More surprisingly, the CTM persisted in its commitment to authoritarianism even after the regime itself began to move in a more democratic direction.[59] The CTM denounced the regime's legalization of leftist parties in 1977, called for the expulsion of Cardena's reformist movement from the ruling party in 1988, protested opposition successes in the general elections held that year, and refused support to opposition parties in the 1990s. Ponte observes that democratization was organized labor's greatest fear.[60] The question is, why?

Both state dependence and aristocracy prove decisive in explaining the political disposition of the CTM. First, with regard to state dependence, the CTM long enjoyed an authoritarian bargain with the Mexican state. In exchange for the CTM's delivery of reliable political support and industrial peace, the regime provided labor with a host of organi-

[58] This analysis draws extensively on Middlebrook (fn. 52); idem, *The Paradox of the Revolution: Labor, State, and Authoritarianism in Mexico* (Baltimore: Johns Hopkins University Press, 1995); Stephen E. Morris, *Political Reformism in Mexico* (Boulder, Colo.: Lynne Rienner, 1995); Collier (fn. 4); and Rueschemeyer, Stephens, and Stephens (fn. 8).

[59] For explanations of the regime's decision to embrace democratic reform, see Morris (fn. 58), 7–31.

[60] Victor Manuel Durand Ponte, "The Confederation of Mexican Workers, the Labor Congress, and the Crisis of Mexico's Social Pact," in Middlebrook (fn. 52), 94, 101–2. See also Middlebrook (fn. 52), 15; Rueschemeyer, Stephens, and Stephens (fn. 8), 217–19; and Middlebrook (fn. 58), 292, 311.

zational and material benefits. These included financial subsidies ar legal concessions (for example, exclusive domain and closed shop), well as profit-sharing schemes, privileged access to social-welfare pr grams, subsidized urban housing and credit, expanded social securit and a nationally defined minimum wage.[61] Although the bargain wi labor was lopsided (the state's superior power, both coercive and a ministrative, enabled it to set the terms of the alliance with labor in self-serving fashion),[62] the organizational and material benefits deli ered by the state *far exceeded what labor could have procured on its own.*

The structurally weak position of organized labor in the Mexica economy made political sponsorship by the state essential for labor's a tainment of basic material and social rights.[63] Many factors (for exan ple, high rates of unemployment, geographic dispersal of the lab force, small worker concentration per firm, low overall levels of unio ization, and global pressures) undercut the structural position of orga ized labor in the Mexican economy.[64] Sponsorship by the sta compensated for some of this weakness. But the logic of state sponso ship hinged on the persistence of an *authoritarian* bargain in Mexi and, more specifically, on the leverage that labor enjoyed as a key gua antor of popular support for the regime. Because democratization pr sented the regime with new ways to forge political legitimacy and buil popular support, it threatened to rob the labor confederation of its k political trump (and its certainty of political sponsorship). As such, proved less than attractive.

In addition to labor's dependence on state sponsorship (and the lin age of state sponsorship to an authoritarian system of rule), labor's di fidence about democratization was also a function of its "aristocrati position in the Mexican economy. Despite decades of developmen minded governments, poverty remained rampant,[65] unemployment an

[61] Collier (fn. 4), 59, 83; Ponte (fn. 60), 100; Lawrence Whitehead, "Mexico's Economic Prospec Implications for State Labor Relations," in Middlebrook (fn. 52), 73.

[62] Middlebrook (fn. 52), 9; Whitehead (fn. 61), 75.

[63] Victoria Murillo, "A Strained Alliance: Continuity and Change in Mexico," David Rockefell Center for Latin American Studies Working Paper no. 96-3 (Cambridge: Harvard University, 199 Middlebrook (fn. 58), 288.

[64] Much of this structural weakness was a consequence of (or at least reinforced by) the late timi of Mexican industrialization. Late timing resulted in an immature industrial sector, still overwheln ingly composed of small, geographically dispersed firms. It led to the importation of turnkey proje from industrialized countries whose capital intensity did little to absorb labor surplus. It meant indu trialization in the context of global capital mobility, reducing the leverage of local labor. For more, s Middlebrook (fn. 58).

[65] In 1989, 23 percent of Mexicans still lived below the poverty line and 7.3 percent lived in "e treme poverty." Sebastian Edwards, *Crisis and Reform in Latin America: From Despair to Hope* (Ne York: Oxford University Press, 1995).

nderemployment remained high,[66] and the informal economy contin-ed to account for a large proportion of the country's economic activ-y.[67] In this context unionized workers in Mexico's formal economy ccupied a privileged economic position, enjoying access to stable em-loyment, as well as nonwage benefits that far exceeded those enjoyed y the vast majority of their compatriots. These benefits included priv-eged access to state-subsidized housing, health care, financial credit, nd retirement funds.[68] This privilege disjoined labor's interests from 1ose of most other poor Mexicans and gave organized labor little in-entive to join forces with other subordinate strata to make the state ac-ountable to mass interests. To the contrary, organized labor had an terest in preserving its special relationship with the state, even if that eant bolstering an authoritarian regime. Thus, the CTM refused to upport Cardenas and other left-leaning candidates in their campaigns f the 1970s and 1980s and continued to support the PRI in an effort to ustain the status quo.[69]

In short, diffidence about democracy has been the mark of lead-ng trade unions in Mexico like the CTM. State dependence and aristo-ratic status go far toward explaining their reluctance to break with uthoritarianism.[70]

.OREA: STATE PERSECUTION, ECONOMIC EXCLUSION, AND DEMOCRATICALLY COMMITTED LABOR[71]

1 contrast to the Mexican case, organized labor in Korea has long been t the forefront of the struggle for democratization. Authoritarian rule 1 Korea was interrupted twice during the postwar era, and on both oc-asions organized labor played an active role in the popular movements

[66] In 1995 urban underemployment reached 25.9 percent and official unemployment figures for ban areas clocked in at 6 percent. See *The Economist: Country Profile, Mexico* (1996–97), 48.

[67] In 1996 the informal sector was the source of jobs for between 20 and 30 percent of the Mexican oor force. *The Economist: Country Profile, Mexico* (1996–97), 10; *Country Report, Mexico, Economist* telligence Unit (1st quarter, 1997), 21.

[68] Middlebrook (fn. 58), 221. Wage concessions won by the CTM generally translated into higher inimum wages that benefited all workers, not just union members. See also Wounter van Ginneken, *'cio-Economic Groups and Income Distribution in Mexico* (London: Croom Helm, 1980), 68–69.

[69] Ponte (fn. 60), 102.

[70] A counterhypothesis suggested by an anonymous reader argues that internal trade union democ-cy, more than state dependence or aristocratic position, might be the better predictor of organized bor's support for democracy. But the fact that democratically inclined trade unions have made pacts ith authoritarianism when handicapped by structural weakness makes me skeptical; Ponte (fn. 60), ', 100; and Enrique de la Garza Toledo, "Independent Trade Unionism in Mexico," in Middlebrook 1. 52), 159, 174. Additional research is necessary to test this counterhypothesis.

[71] This analysis draws extensively on Eckert (fn. 30, 1990); Ogle (fn. 30); Koo (fn. 45); Frederic eyo, *Beneath the Miracle: Labor Subordination in the New Asian Industrialism* (Berkeley: University of alifornia Press, 1989).

agitating for reform.[72] In 1960 workers followed student-initiated protest, organizing scores of strikes and demonstrations that called for an end to corrupt government.[73] Similarly, during the 1980s workers joined forces with students, intellectuals, and church activists to pressure the regime to reform.[74] In contrast to Mexico's CTM, organized labor staunchly allied itself with the cause of democratization in Korea. What explains this divergence in labor's behavior?

The degree of labor's state dependence and aristocratic privilege prove important. Overall, labor did not enjoy a dependent relationship with the state in Korea. Labor's relative independence was a consequence of both state policy and labor's own market position. With regard to the state's posture toward labor, Korea stands in stark contrast to Mexico. In Mexico the regime's "revolutionary" origins inclined it toward alliance with the working class, and an import-substituting strategy of industrialization (that persisted well into the 1970s) provided it with sufficient economic space to accommodate labor's demands. In Korea, by contrast, the regime's emergence in a cold war context and its experience of civil war with the North made it extremely sensitive to the threat of communist overthrow. This spelled hostility toward organized labor, which was perceived as a potential vehicle for communist infiltration. In addition, the regime's early embrace of export-oriented industrialization put a premium on low-cost, quiescent labor, setting the regime at odds with organized labor.

Given these conditions, ruthless repression of labor (not cozy corporatist alliance) emerged as the defining mark of the regime. From the goon squads who hounded labor activists during the U.S. occupation to the legal emptying of the labor movement during Syngman Rhee's reign to the mass arrests of union militants under Park Chung Hee to the espousal of torture and "purification camps" under Chun Doo-Hwan, every postwar government in Korea repressed labor, often brutally.[75] Labor therefore had little incentive to side with the regime and every reason to push for reform that would make the state more responsive to mass interest.

Lack of state support thus gave labor both the independence and the interest to embrace the cause of democratization. But labor's inde-

[72] For historical details, see Eckert (fn. 30, 1990), 352–56; and Hamilton and Kim (fn. 30), 118–21.

[73] Ogle (fn. 30), 13–16.

[74] Eckert (fn. 30, 1990), 380; Jeonge Taik Lee, "Dynamics of Labor Control and Labor Protest in the Process of EOI in South Korea," *Asian Perspective* 12 (Spring 1988), 149; Koo (fn. 30), 39; Ogle (fn. 30), 116.

[75] For details of this repression, see Ogle (fn. 30); Koo (fn. 30); Lee (fn. 74); and Deyo (fn. 71).

pendence was also fostered by its market position. In contrast to the Mexican case, where high levels of unemployment, low levels of industrialization, and low skill levels spelled a structurally weak position for labor (and hence dependence on state propping), labor in Korea enjoyed a relatively strong structural position. Korea's rapid growth rate and the dramatic expansion of industry led to an increasingly tight labor market that strengthened labor's negotiating muscle.[76] In addition, the regime's shift into heavy industry reinforced labor's position by creating enormous industrial sites that concentrated workers and facilitated trade-union organizing. Furthermore, by focusing on industries that relied on skilled labor, this shift in strategy insulated labor from competition with the poorest of the poor in the global proletariat.[77] In short, Korea's development path favored labor with independent structural power, and this meant labor could look to itself rather than to state propping to advance its interests. This also freed labor to embrace political campaigns not favored by state elites.

Finally, Korean labor also scored low in terms of aristocracy. For all its growing structural strength, labor hardly enjoyed a privileged economic position in Korean society. Wages remained deplorably low, lagging far behind productivity gains and substantially trailing wages in many service sectors.[78] Working conditions were appalling.[79] And working hours were interminable.[80] This hardship was especially difficult for workers, given the material progress made by so many other sectors in society.[81] Workers "came to see themselves as the principal victims of economic development,"[82] not as the beneficiaries of aristocratic privilege. Worse still, the harsh treatment of workers was the deliberate intent of the regime, which saw the extreme exploitation of labor as the cornerstone of its development strategy. In contrast to the situation in Mexico, organized labor in Korea had no special, politically

[76] Koo (fn. 45), 1030; Deyo (fn. 71), 24.

[77] Wage comparison was made with workers in Europe and the U.S., not Bangladesh.

[78] Throughout the 1970s only 10 percent of workers in manufacturing and mining earned incomes equal to the minimum living standard set by the government. Only 50 percent made even half of that standard. Ogle (fn. 30), 76. See also Kong (fn. 44), 240. For figures on the failure of wages to keep pace with gains in productivity, see Hak-Kyu Sohn, *Authoritarianism and Opposition in South Korea* (London: Routledge, 1989), 234, n. 81. For wage differentials between industrial workers and workers in the service sector and agriculture, see Young-Ki Park, *Labor and Industrial Relations in Korea* (Seoul: Sogang University Press, 1979), 102; and Young-Bum Park, *Labor in Korea* (Seoul: Korea Labor Institute, 1993), 59.

[79] Manufacturers routinely ignored even the most basic health and safety regulations; Ogle (fn. 30), 77.

[80] Manufacturing workers averaged the longest workweek in the world, officially clocking in at fifty-four hours per week (though some argue that sixty hours per week was a more common average). Ogle (fn. 30); Deyo (fn. 71), 98.

[81] Kong (fn. 44), 226–37.

[82] Lee (fn. 74), 144.

mediated privilege to conserve and therefore every reason to support democratization of the political system.

In the Korean case low scores on state dependence and aristocratic position help explain organized labor's interest in and capacity for embracing democratization.

Corroboration from Zambia, Egypt, and Tunisia

The importance of state dependence and aristocratic position for shaping labor's commitment to democracy is further corroborated by evidence from Zambia, Egypt, and Tunisia. The Zambian case confirms that low scores on both these variables incline labor toward championing democratization.[83] The Zambian trade union movement (ZCTU), long one of the strongest in sub-Saharan Africa, was based primarily in copper mining. This sector provided the lion's share of the country's foreign exchange and government revenues. During the 1980s, however, a steep decline in international copper prices plunged the country into economic crisis. Workers experienced a huge erosion in wages, eliminating any vestige of aristocratic privilege.[84] The state, moreover, did nothing to insulate labor from the crisis but rather embarked on an energetic campaign of labor repression.[85] In this context of low aristocracy and negative state sponsorship, Zambia's labor movement emerged as one of the most enthusiastic campaigners for democratization. The ZCTU threw its support behind the Movement for Multi-Party Democracy, organizing strikes, mobilizing an extensive network of trade-union committees, and providing key leadership for the cause.[86] Like Korea, Zambia provides an example of a trade-union movement that once cut loose from the moorings of state support and exposed to harsh economic conditions emerges as a major champion of democratic reform. The Zambian case is also useful because it helps eliminate the rival hypothesis that links enthusiasm for democracy with economic growth and prosperity. In Zambia labor's activism on behalf of democratization came precisely at a time of great economic crisis. Compare this with the Korean case, where labor's agitation for democratization developed in a context of economic prosperity.

[83] This analysis draws extensively on Paschal Mihyo, "Against Overwhelming Odds: The Zambian Trade Union Movment," in Henk Thomas, ed., *Globalization and Third World Trade Unions* (London Zed Books, 1995), 201–14.

[84] Low- and middle-income workers saw their wages decline by an average of 55 percent during the 1980s. See Mihyo (fn. 83), 203.

[85] Ibid., 208.

[86] Ibid., 201. For more, see L. Rakner, *Trade Unions in Processes of Democratization: A Study of Party Labour Relations in Zambia* (Bergen: Michelsen Institute, 1992).

In contrast to the Zambian case, the experience of organized labor in Tunisia replicates the pattern of the CTM in Mexico and shows that high scores on state dependence and aristocratic position can spell labor diffidence about democratization.[87] Organized labor (the UGTT) in Tunisia long enjoyed a collaborative, if troubled, relationship with the state. The state favored the trade-union movement with financial subsidies, a relatively generous legal context, and a host of wage and nonwage benefits for its members, in exchange for labor's political alliance.[88] These benefits privileged organized workers vis-à-vis workers in the country's massive and unorganized informal sector, and they far exceeded what labor would have commanded unassisted. But though the state favored organized labor in this way, it brooked no hint of political independence on the part of the UGTT and answered even the most minor challenge to state hegemony with brutal and unrelenting repression. As a result, the trade-union movement has distanced itself from the country's fledgling movements for democracy, denying them both financial and symbolic support.[89] In addition, the UGTT has consistently thrown its support to President Ben Ali, despite his regime's increasing repressiveness and authoritarianism. Labor's desire not to upset the apple cart of state patronage and privilege has prevented it from embracing the campaign for democratization.

The cases cited above have shown a fair degree of collinearity between state dependence and aristocratic position. In Egypt, however, the state dependence of the labor movement remains *high* but its aristocrat privilege is *declining*.[90] The labor movement was long party to a corporatist bargain with the Egyptian regime, exchanging worker restraint for essential material and organizational benefits from the

[87] This section draws extensively on Eva Bellin, "Stalled Democracy: Capitalist Industrialization and the Paradox of State Sponsorship in Tunisia, the Middle East, and Beyond" (Book manuscript), chaps. 4, 5.

[88] For example, throughout the 1990s the regime intervened in national wage negotiations to bolster the position of labor and guarantee wage gains that, while not dramatic, far exceeded what labor could command on its own. For more, see Bellin (fn. 87); and Christopher Alexander, "State, Labor, and the New Global Economy in Tunisia," in Dirk Vandewalle, *North Africa: Development and Reform in a Changing Global Economy* (New York: St. Martin's Press, 1996).

[89] For example, the UGTT has refused to publish any of the reports or declarations of Tunisia's Human Rights League (the crusading force for civil liberties in the country) in the trade union newspaper; nor has it ever publicly endorsed the league's work.

[90] This analysis draws extensively on Marsha Pripstein Posusney, *Labor and the State in Egypt* (New York: Columbia University Press, 1997); Joel Beinin and Zachary Lockman, *Workers on the Nile* (Princeton: Princeton University Press, 1987); Robert Bianchi, "The Corporatization of the Egyptian Labor Movement," *Middle East Journal* 40 (Summer 1986); idem, *Unruly Corporatism: Associational Life in Twentieth-Century Egypt* (New York: Oxford University Press, 1989); Ellis Goldberg, "The Foundations of State-Labor Relations in Contemporary Egypt," *Comparative Politics* 24 (January 1992).

state.[91] Both sorts of benefits spelled aristocracy and dependence for organized labor, and this politically mediated privilege wedded the trade-union movement to the authoritarian status quo, making it unreceptive to democratic reform. Thus, when Sadat began to move Egypt toward a multiparty system in the early 1970s, trade-union leaders vigorously condemned the plan because it threatened to reduce the influence they had enjoyed under the country's single-party system.[92]

Recently, however, organized labor has begun to shift its stance on democratization. In the last decade international pressure to undertake structural adjustment has forced the regime to retreat from its historic bargain with labor. Specifically, the regime has begun to reform the labor code in ways that will ax organized labor's most treasured advantage (job security), and it has endorsed privatization schemes that will contract the sphere of labor's political protection.[93] With the essence of labor's aristocratic privilege now under attack, trade-union leaders have begun to express support for democratization.[94] Persistent dependence makes the labor movement hesitant about dramatically endorsing political reform (for example, it has not formed an independent labor party to contest the ruling party), but declining aristocracy (and hence the declining value of the authoritarian bargain) makes labor receptive to reform, so long as it comes at someone else's initiative. All told, organized labor has shifted from candid negativism to an ambivalent attitude toward democratization, with a declining score on aristocratic privilege the best explanation for this change of heart.

CONCLUSION

The case studies presented in this article demonstrate that capital and labor are contingent, not consistent, democrats. This contingency, moreover, is not random. Support for democratization turns on whether capital and labor see their economic interests served by the authoritarian state. This, in turn, is shaped by two key factors for each social force. For capital, democratic enthusiasm hinges on its level of state dependence and fear of social unrest. For labor, democratic enthusiasm hinges on its level of state dependence and aristocratic position in soci-

[91] These included financial subsidies to the union, prestigious political positions for union leaders, and nonwage benefits for workers such as job security, social security benefits, and generous leave policy (maternity).

[92] See Bianchi (fn. 90, 1986) 438; and idem (fn. 90, 1989), 138.

[93] See Marsha Pripstein, "Egypt's New Labor Law Removes Worker Provisions," *Middle East Report* (May–August 1995), 52–53.

[94] Bianchi anticipated this development as early as the mid-1980s. See Bianchi (fn. 90, 1986), 443.

ety. The relationship is an inverse one, with higher values of dependency, fear, and aristocracy translating into reduced enthusiasm for democratic reform. In many late-developing countries a number of factors—including extensive state sponsorship, the structural weakness of social forces, pervasive poverty, and the evolution of democratic discourse—have led capital and labor to ally with authoritarian states rather than championing democratization. This experience diverges from that of the earlier industrializers—or at least from the historical myth of the early industrializers—and has political consequences that challenge the expectations of classic liberal and Marxist analysts.

At the same time, capital and labor's alliance with authoritarianism is not carved in stone. The political disposition of capital and labor is governed by interest. As political and economic conditions change, interests may change and alliances may be recalculated. Our case studies suggest some of the conditions that may prompt such realignment. The logic of international economic integration may force the state to reduce its sponsorship of social forces (Korea, Egypt). Or robust growth may eliminate mass poverty and the pervasive sense of fear within the propertied class (Korea). Under these conditions capital and/or labor may perceive democratization in a new light and choose to embrace it. The intertemporal variation found within cases like Korea, Brazil, and Egypt makes this possibility clear.

Predicting societal pressure for democratization, then, turns on the analysis of these variables. Capital and labor may well be champions of democracy, but for contingent reasons and not by universal dictum. The particular conditions of late development may dampen social forces' enthusiasm for democratization. But contingency also spells the possibility of democratic enthusiasm as conditions in late-developing countries change.

THE RELIGIOUS ROOTS OF MODERN INTERNATIONAL RELATIONS

By DANIEL PHILPOTT*

SOME recent trends in international relations—the United Nations' imprimatur for intervention in the politics of broken, war-torn, malnourished, and dictatorial states, the formation of the European Union in the 1991 Maastricht Treaty—depart notably from what international relations scholars refer to, out of common grammar and historical consensus, as the Westphalian system of sovereign states.

When a political order ruptures, its rival inhabitants will seek out its origins—conservatives, to fortify the system's pedigree; revolutionaries, to reveal its flawed foundations; scholars, to descry the sorts of winds that brought the order and the sorts that might carry it away. This article takes on the scholar's task—if the crumbling order is Westphalian, then how did it come to pass?—and it proposes that a crucial spring of our state system was the Protestant Reformation. Religious ideas, it is argued, are at the root of modern international relations. The Reformation was not Westphalia's sole cause; long-term material trends contributed, too. But if ideas did not act alone, they were yet indispensable: no Reformation, no Westphalia.

The claim is a counterfactual one. Had the Reformation not occurred, a system of sovereign states would not have arrived, at least not in the same form or in the same era. Skeptics imagine an alternative world in which the Reformation never happened, yet one in which

* Previous versions of this article were delivered at the John M. Olin Institute for Strategic Studies, Harvard University, 1994, the Watson Institute, Brown University, 1995, the Center of International Studies, Princeton University, 1996, the annual meeting of the American Political Science Association, San Francisco, 1996, and the Institute on Global Conflict and Cooperation, La Jolla, California, May 1997. For providing fellowship support and facilities to conduct the research for the article, I thank the Center for International Affairs, Harvard University, the John M. Olin Institute for Strategic Studies, Harvard University, the Research Program on International Security at the Center of International Studies, Princeton University, and the Erasmus Institute, University of Notre Dame. For helpful comments on the manuscript, I thank Aaron Belkin, Vikram Chand, Jarat Chopra, Benjamin Cohen, Colin Elman, Peter Gourevitch, Rodney Bruce Hall, Bryan Hehir, Stanley Hoffmann, Robert Keohane, Elizabeth Kier, Stephen Krasner, Friedrich Kratochwil, Andrew Moravcsik, Henry Nau, John Owen, Timothy Shah, Daniel Thomas, and Barbara Walter.

Westphalia nevertheless arrived when and how it actually did and was brought by the same forces that social scientists most commonly identify to explain the rise of individual states—shifts in economic and organizational structures, in trade, class, societal coalitions, wealth, technology, military might, the institutions of domestic coercion, and the international balance of power. But such a world, it is argued here, could not have yielded Westphalia.

The Reformation's indispensability emerges most saliently through the following correlation: those polities that experienced a Reformation crisis were the same ones that adopted an interest in Westphalia; those that saw no such crisis did not. In plumbing the causal logic behind this correlation, I argue that the intrinsic content of Protestantism itself points to sovereignty. Then, considering the diverse historical pathways taken by Germany and France, I show how Protestant ideas led to interests in Westphalia. I also argue that such strong correlations do not obtain in the skeptic's world; that is, polities favoring Westphalia failed to acquire more powerful state institutions, arms, and wealth prior to their interest in Westphalia, while there were relatively powerful and wealthy polities that opposed Westphalia. Indeed, much of the material growth and decline that helped to realize Westphalia—the "military revolution," for instance—was itself due in important part to the Reformation. Geographical and temporal correlations, combined with the historical accounts of how separate polities arrived at an interest in a system of sovereign states, together make the case for the efficacy of the Protestant Reformation.

The argument presented here contributes to international relations scholarship in three ways. First, it accounts for the origin of Westphalia, of the very authority structure of the system of sovereign states. International relations scholars have long granted that a state system exists and have sought to theorize its laws and patterns of war, peace, and commerce. Recently, though, several scholars in the field have come to emphasize the "constructed," historically contingent character of the international system itself—an emphasis echoed in the article. A few of these constructivists have even begun to chart the genesis of the system's very political authority.[1] It is these rules that define au-

[1] These include Friedrich Kratochwil, "Of Systems, Boundaries, and Territoriality: An Inquiry into ie Formation of the States System," *World Politics* 39 (October 1986); John Gerard Ruggie, "Territoality and Beyond: Problematizing Modernity in International Relations," *International Organization* 7 (Winter 1993); Jens Bartelson, *A Genealogy of Sovereignty* (Cambridge: Cambridge University ress, 1995); and Rodney Bruce Hall, *National Collective Identity: Social Constructs and International ystems* (New York: Columbia University Press, 1999). On the works of the constructivists more genally, see Alexander Wendt, "The Agent-Structure Problem in International Relations Theory,"

thority—or sovereignty—in the international system whose origins want to explain.

A second contribution is a corresponding view of what fueled this epochal change—the kinetic energy of ideas. Most scholars of the origins of the state, including those who have addressed the states system, have privileged material changes, and their work has yielded a vast new understanding of how the state and the state system came to be.[2] But few of them have looked to ideas, still less to religious ideas, as a crucial source. Should it prove efficacious, the Reformation would warrant recognition as a kind of historical cause that merits more attention in the international relations literature.

Yet the very premise of the argument that Westphalia is the origin of modern international relations has itself recently come under attack, prominently from Stephen Krasner.[3] The third contribution, then, is an updated and qualified defense of this conventional wisdom. As this premise undergirds the first two contributions, its defense is the first task of the essay.

WESTPHALIA AS ORIGIN

In his classic 1948 article, the international legal scholar Leo Gross calls Westphalia "the majestic portal which leads from the old world

International Organization 41 (Summer 1987); idem, "Anarchy Is What States Make of It: The Social Construction of Power Politics," *International Organization* 46 (Spring 1992); idem, "Collective Identity Formation and the International State," *American Political Science Review* 88 (June 1994); John Gerard Ruggie, "What Makes the World Hang Together? Neo-utilitarianism and the Social Constructivist Challenge," *International Organization* 52 (Autumn 1998); Martha Finnemore and Kathryn Sikkink, "International Norm Dynamics and Political Change," *International Organization* 52 (Autumn 1998); Jeffrey T. Checkel, "The Constructivist Turn in International Relations Theory," *World Politics* 50 (January 1998). For works on sovereignty and its history, see John Gerard Ruggie, "Continuity and Transformation in the World Polity: Toward a Neorealist Synthesis," in Robert O. Keohane, ed., *Neorealism and Its Critics* (New York: Columbia University Press, 1986); Ruggie (fn. 1, 1993); Kratochwil (fn. 1); Thomas J. Bierksteker and Cynthia Weber, eds., *State Sovereignty as Social Construct* (Cambridge: Cambridge University Press, 1996); Bartelson (fn. 1); Daniel Philpott, "Sovereignty: An Introduction and Brief History," *Journal of International Affairs* 48 (Winter 1995); Stephen D. Krasner, "Sovereignty: An Institutional Perspective," *Comparative Political Studies* 21 (April 1988); F. H. Hinsley, *Sovereignty*, 2d ed. (Cambridge: Cambridge University Press, 1986); Alan James, *Sovereign Statehood* (London: Allen and Unwin, 1986); Hendrik Spruyt, *The Sovereign State and Its Competitors* (Princeton: Princeton University Press, 1994); Janice Thomson, "State Sovereignty in International Relations: Bridging the Gap between Theory and Empirical Research," *International Studies Quarterly* 39 (June 1995); Michael Ross Fowler and Julie Marie Bunck, *Law, Power, and the Sovereign State* (University Park: Penn State Press, 1995); Barry Buzan, "The Idea of 'International System': Theory Meets History," *International Political Science Review* 15 (July 1994); Stephen Krasner, *Sovereignty: Organized Hypocrisy* (Princeton: Princeton University Press, 1999).

[2] I refer to these works in fnn. 47–50.

[3] Krasner, "Westphalia and All That," in Judith Goldstein and Robert O. Keohane, eds., *Ideas and Foreign Policy: Beliefs, Institutions, and Political Change* (Ithaca, N.Y.: Cornell University Press, 1993).

into the new world." Hans Morgenthau and other leading political scientists echoed Gross, rendering Westphalia metonymy for the modern international system.[4] Why the audacious ascription? Westphalia, such scholars hold, defined an international scheme of constitutional authority that centered upon sovereign states—polities with supreme authority located within a bordered territory.[5]

Recent iconoclasm, however, challenges Westphalia's "majesty." Krasner, for instance, argues that before Westphalia, Europe was substantially modern—states with sovereign privileges already existed—and that after Westphalia, Europe was persistently medieval: the Holy Roman Empire and several compromises of sovereignty, including minority agreements for religion, remained. On the trek to modernity, then, Westphalia was not a continental divide but just another hill.[6] The iconoclasm is strong, not easily dismissed.

Westphalia's prestige can be rescued, but only with a new defense that is more subtle and qualified than the conventional wisdom. The nub of the Westphalian claim: before 1648, as long as the Thirty Years' War was still smoldering, political authority in Europe was essentially incompatible with sovereign statehood, whereas afterward sovereignty generally prevailed. I argue, however, that Westphalia signals the consolidation, not the creation ex nihilo, of the modern system.[7] It was not an instant metamorphosis, as elements of sovereign statehood had indeed been accumulating for three centuries. Even modernity's victory after Westphalia must be qualified, for some medieval anomalies persisted. But consummate fissures in history are rare, and Westphalia is as clean as historical faults come.

Politics after Westphalia appears sharpest in relief against the Europe of the High Middle Ages, between the eleventh and thirteen centuries. In the *Respublica Christiana*, there was no supreme authority

[4] Leo Gross, "The Peace of Westphalia, 1648–1948," *American Journal of International Law* 42 (January 1948); Hans Morgenthau, *Politics among Nations: The Struggle for Power and Peace*, 6th ed. (New York: Alfred A. Knopf, 1985), 328–30.

[5] Sharing my general definition of sovereignty are, among others, Spruyt (fn. 1), 34–36; Ruggie (fn. 1, 1993), 148–52; James (fn. 1); Hinsley (fn. 1), 158; Morgenthau (fn. 4), 32–38; Robert O. Keohane, "Sovereignty, Interdependence, and International Institutions," in Linda B. Miller and Michael Joseph Smith, eds., *Ideas and Ideals: Essays on Politics in Honor of Stanley Hoffmann* (Boulder, Colo.: Westview Press, 1993), 92–93; J. L. Brierly, *The Law of Nations* (New York: Oxford University Press, 1963), 13; Kenneth Waltz, *Theory of International Politics* (Lexington, Mass.: Addison-Wesley, 1979), 96. For skeptics of attempts to define sovereignty, see Bartelson (fn. 1); E. H. Carr, *The Twenty Years' Crisis* (New York: Harper and Row, 1964); Stanley Benn, "Sovereignty," *Encyclopedia of Philosophy* 7 (1967); L. Oppenheim, *International Law*, vol. 1 (London: Longmans, Green, 1905); Richard Falk, "Sovereignty," in *Oxford Companion to Politics of the World* (Oxford: Oxford University Press, 1993).

[6] See especially Krasner (fn. 3), 235–64.

[7] For a similar view, see Martin Wight, *Systems of States* (London: Leicester, 1977); and Charles Tilly, *Coercion, Capital, and European States, AD 990–1992* (Oxford: Basil Blackwell, 1992), 161–83.

within a territory, manifestly no sovereignty.[8] *This dispersed au of course, did not last forever. The European landscape came gra to be filled with sovereign states over the next three and a hal turies. By the eve of the Reformation in 1517, monarchs in Bri France, and Sweden had established supremacy over the church* other territorial rivals, while in Italy, a small system of sovereign st had survived for a century.

But Europe's metamorphosis into modernity was far from finish At roughly this time the Italian states system was conquered fr without, and Charles V became both king of Spain and emperor of Holy Roman Empire, acquiring substantial, but not sovereign, pow in lands extending from the Netherlands to Austria. The church w still temporally formidable, too, with its vast holdings of land a maj source of wealth and power. Within the empire alone the church he one-third of the land overall, including one-fourth of the property i the cities.[9] Its archbishops and bishops occupied political offices. Th church partially governed education, raised considerable revenue helped administer poor relief and other public affairs, and most of al often with the cooperation of temporal rulers, monitored the religiou allegiance of its members. Having sworn allegiance to the pope, to up- holding the faith, Charles V served as the temporal enforcer of ecclesi- astical orthodoxy, the commander of Catholic armies.[10]

[8] Susan Reynolds, *Kingdoms and Communities in Western Europe, 900–1300* (Oxford: Clarendo Press, 1984); Gianfranco Poggi, *The Development of the Modern State: A Sociological Introduction* (Stan ford, Calif.: Stanford University Press, 1978); Georges Duby, *La societé aux XI^e et XII^e siècles dans la r gion maconnaise* (Paris: Ecole Pratique des Hautes Etudes, 1953). For a challenge to the consensu among medieval historians that the Middle Ages generally lacked sovereignty, see Markus Fische "Feudal Europe, 800–1300: Communal Discourse and Conflictual Practices," *International Organiza tion* 46 (Spring 1992). For responses to Fischer, see Rodney Bruce Hall and Friedrich Kratochwi "Medieval Tales: Neorealist 'Science' and the Abuse of History," *International Organization* 47 (Sum mer 1993); Rodney Bruce Hall, "Moral Authority as a Power Resource," *International Organization* 5 (Autumn 1997). On the nature of political authority during the Middle Ages, see J. R. Strayer, *Th Medieval Origins of the Modern State* (Princeton: Princeton University Press, 1970); J. R. Strayer an Dana Munro, *Middle Ages, 395–1500* (New York: Appleton-Century-Crofts, 1959); Brian Tierne *Crisis of Church and State* (Englewood Cliffs, N.J.: Prentice-Hall, 1964); John H. Mundy, *Europe in t High Middle Ages, 1150–1309* (New York: Basic Books, 1973); Walter Ullman, *Principles of Gover ment and Politics in the Middle Ages* (New York: Barnes and Noble, 1974); Ernst Kantorowicz, *Th King's Two Bodies* (Princeton: Princeton University Press, 1957); Ruggie (fn. 1, 1986), 141–48; Spru (fn. 1), 34–36; Michael Wilks, *The Problem of Sovereignty in the Middle Ages* (Cambridge, U.K.: At th University Press, 1964).

[9] Steven Ozment, *The Age of Reform, 1220–1550: An Intellectual and Religious History of Late Me dieval and Reformation Europe* (New Haven: Yale University Press, 1980), 48; Robert A. Kann, *A Hi tory of the Habsburg Empire* (Berkeley: University of California Press, 1974), 1–24; H. G Koenigsberger, *The Habsburgs and Europe, 1516–1560* (Ithaca, N.Y.: Cornell University Press, 197 Jean Bérenger, *History of the Habsburg Empire* (London: Longman, 1994).

[10] On the implosion of the Italian states system, see Wight (fn. 7); Tilly (fn. 7), 77–78. On Char V's powers, see Kann (fn. 9), 1–24; Koenigsberger (fn. 9); Bérenger (fn. 9); Jonathan Israel, *The Du*

ater in the century sovereignty made gains, but these were frail and ersible.[11] Most importantly, the 1555 Peace of Augsburg allowed rman princes to enforce their own faith within their territory ac- ding to the formula *cuius regio, euis religio* (whose the region, his the igion).[12] But Augsburg did not last. Endless clauses and mutual dis- isfaction with the agreement brought war, expanding eventually into titanic Thirty Years' War. Not until the close of this war did a pro- iption of such intervention become accepted and respected and prac- ced—not, indeed, until 1648.

What, then, was this Peace of Westphalia of 1648? Although the omponent treaties of the settlement, Münster and Osnabrück, never xplicitly mention sovereignty, an analysis of the treaties' text, of the in- ntions of its framers, and of state practice following Westphalia shows ie settlement to be a fulcrum, not just a milepost, in the evolution of ie sovereign states system. This is apparent in two respects: in the re- pective status of states and imperial institutions and in the matter of ligion.[13]

Although Westphalia was officially a new set of laws for the Holy oman Empire, which lasted until 1806, in effect the settlement be- ieathed to states absolute sovereign authority.[14] In the wake of West- ialia states became the chief form of polity in Europe and faced no rious rival in the Holy Roman Empire—this is the heart of the claim out historical change. The United Provinces and the Swiss Confed- ation gained effective independence. The German states regained

---

public: Its Rise, Greatness, and Fall, 1477–1806 (Oxford: Oxford University Press, 1995), 9–40; Pieter yl, *The Revolt of the Netherlands, 1555–1609* (London: Williams and Norgate, 1932); Gerald auss, *Law, Resistance, and the State: The Opposition to Roman Law in Reformation Germany* (Prince- i: Princeton University Press, 1986); Hajo Holborn, *A History of Modern Germany*, vol. 1 (New York: fred A. Knopf, 1959); John Gagliardo, *Germany under the Old Regime, 1600–1790* (London: Long- ins, 1991); Geoffrey Barraclough, *The Origins of Modern Germany* (Oxford: Basil Blackwell, 1947).

[11] Koenigsberger (fn. 9); Barraclough (fn. 10), 355–405; Holborn (fn. 10), 284–338.

[12] Holborn (fn. 10), 243–46; Barraclough (fn. 10), 371.

[13] Andreas Osiander, *The States System of Europe, 1640–1990* (Oxford: Clarendon Press, 1994), 72. r the text of the Treaty of Münster, see Fred Israel, ed., *Major Peace Treaties in Modern History, 48–1967* (New York: Chelsea House Publishers, 1967); for the Treaty of Osnabrück, see Clive Parry, , *The Consolidated Treaty Series* (Dobbs Ferry, N.Y.: Oceana Publications, 1981). On the settlement general, see Fritz Dickmann, "Rechtsgedanke und Machtpolitik bei Richelieu: Studienen neu en- kten Quellen," *Historische Zeitschrift* 196 (April 1963); idem, *Der Westphaelische Frieden* (Münster: chendorff, 1965); Fritz Dickmann et al., eds., *Acta Pacis Westphalicae* (Münster: Aschendorff Ver- sbuchhandlung, 1962); George Pages, *The Thirty Years War*, trans. David Maland (New York: rper and Row, 1970); Geoffrey Parker, *The Thirty Years' War* (London: Routledge, Kegan, and Paul, 97); J. V. Polisensky, *The Thirty Years' War*, trans. Robert Evans (Berkeley: University of California ess, 1971); and T. K. Rabb, ed., *The Thirty Years' War: Problems of Motive, Extent, and Effect* (Boston: C. Heath and Company, 1964).

[14] On the settlement in general , see Dickmann et al. (fn. 13); Dickmann (fn. 13, 1965 and 1963); ges (fn. 13); Parker (fn. 13); Polisensky (fn. 13); and Rabb (fn. 13).

their "ancient rights" against the empire and acquired the right to fo alliances outside the empire.[15] The communications between diplom at Westphalia are laced with references to state autonomy, the equali of states, an equilibrium of states, and even an early version of collecti security—all notions that are unintelligible apart from a sovereig states system.[16] Following Westphalia, states became virtually uninhib ited in their internal authority. Imperial institutions functioned as forum for German cooperation, but they did not abridge sovereignty The emperor himself retained significant constitutional powers onl because he and his successors were Habsburgs, but the Habsburg themselves were effectively reduced to just another sovereign, withi their central European lands.[17]

Politics and religion provided the other major context in which Westphalia effectively ushered in state sovereignty. Here, too, the text of the treaties does not explicitly establish sovereignty—at least not in the exact form of Augsburg in 1555, which granted German princes authority over religion in their lands. Instead, princes agreed to allow designated proportions of Catholic, Lutherans, and Calvinists to live and practice their faiths in their territories, and to refrain from attempting to convert one another's subjects.[18]

But the crucial question for sovereignty is not whether princes signed agreements regarding their internal affairs; modern states, after all, frequently agree to regulate prices or pollution within their borders without compromising their sovereignty. Rather, the issue is whether any institution outside the state was vested with the constitutional authority to execute, judge, or legislate about these agreements about religion. Here, the text of the treaties calls for arbitration of religious disputes, but through compromise, not majority decision, thus leaving

[15] Israel (fn. 13), 27; Osiander (fn. 13), 46–47.

[16] On the views of the diplomats, see Osiander (fn. 13), 27, 41, 77–89; William F. Church, *Richelieu and Reason of State* (Princeton: Princeton University Press. 1972), 283–349; Michael Roberts, *Gustavus Adolphus*, 2 vols. (London: Longmans, Green, 1958); idem, *Essays in Swedish History* (London: Weidenfeld and Nicolson, 1967), 82–110; Carl J. Burckhardt, *Richelieu and His Age* (New York: Harcourt, Brace, and World, 1970); Robert Knecht, *Richelieu* (London: Longman, 1991); Armand-Jean du Plessis Richelieu, *The Political Testament of Cardinal Richelieu: The Significant Chapters and Supporting Selections*, trans. Henry Bertram Hill (Madison: University of Wisconsin Press, 1981).

[17] Exceptional were the imperial circles, regional organizations of princely states that revived in the late seventeenth century and whose members would at times combine their military forces under a single command. But this exception was limited: the strong form of military cooperation was limited largely to the circles of Franconia and Swabia, composed only of small German states, and it lasted only from 1697 to 1714, when states defended themselves against the invasions of Louis XIV. See Roger Wines, "Imperial Circles, Princely Diplomacy, and Imperial Reform, 1681–1714," *Journal of Modern History* 39 (March 1967), 27–29. More generally, see Kann (fn. 9), 52, 54; Osiander (fn. 13), 46; Barraclough (fn. 10), 381–87; and Gagliardo (fn. 10).

[18] Osiander (fn. 13), 40; Holborn (fn. 10), 368–69.

the sovereign right of assent virtually intact.[19] But given the outlooks of contemporary statespersons and diplomats and considering subsequent practice, these procedures would mean little. Historians of the period commonly agree that even in the late stages of the Thirty Years' War, political leaders had lost their religious zeal and ceased to contest religion as a political affair.[20] There is also little record of any religious matter being adjudicated through imperial institutions. Finally, for over a century and a half after Westphalia, unlike after Augsburg, neither princes nor the emperor intervened to contest religion within another prince's territory. Even though future European treaties compromised sovereignty on behalf of religious minorities, never again would intervention for religion play anywhere near the role that it did between 1517 and 1648. Although religion was occasionally a partial contributor to war between states, in general, religion ceased to be a casus belli.[21]

It is in the context of the political history of the mid-sixteenth to mid-seventeenth century that one best discerns the enormous historical significance of state sovereignty. Religious war and its attendant abridgments of sovereignty were the chief sources of contention in European politics for over a century, but after Westphalia this centrally defining form of conflict ceased. International pronouncements and practices then evolved to legitimate and deepen Westphalia's achievement. First in Europe and then gradually elsewhere states became the only polity to exercise significant executive, legislative, and judicial power within a territory. Over the coming century philosophers like Emmeric Vattel and Christian Wolff came to recognize this reality in articulating doctrines of international law and state legal equality; states themselves solidified it by creating diplomatic offices and practices appropriate to a system of states. European states also developed standards of recognition by which polities would be accepted as members of the system only if they displayed essential attributes of statehood.[22] Finally, states came to espouse nonintervention as a cornerstone of international diplomacy.[23] All of this is what makes Westphalia revolutionary.

[19] Osiander (fn. 13), 40–42.

[20] Parker (fn. 13), 196–97; Holborn (fn. 10); Barraclough (fn. 10); Rabb (fn. 13); David Maland, *Europe in the Seventeenth Century* (London: Macmillan, 1966); Kalevi Holsti, *Peace and War: Armed Conflicts and International Order, 1648–1989* (Cambridge: Cambridge University Press, 1991), 46–59.

[21] Holsti (fn. 20), 46–59; Osiander (fn. 13), 49; Barraclough (fn. 10), 381–87.

[22] On recognition practices, see H. Lauterpacht, *Recognition in International Law* (Cambridge, U.K.: At the University Press, 1947).

[23] See R. J. Vincent, *Non-Intervention in International Order* (Princeton: Princeton University Press, 1974).

The Origin of Westphalia

But if Westphalia was an origin, what was the origin of Westphalia? That the Reformation elicited Westphalia is a claim about a contradiction between the existing order and iconoclastic ideas. Such a contradiction must have developed between 1517, the beginning of the Reformation, and 1648, the year of Westphalia. The fact that in 1517 Europe was ebbing away from sovereignty as Charles V consolidated his lands intimates that during this period some novel source of gravity arose to reverse history's tide back toward sovereignty. We can locate this reversal at the strewn dates when several European polities expressed their initial interest in a system of states. An account that looks to ideas will note that in many of these cases this interest appeared roughly when Protestantism did. Even skeptical accounts cannot deny this coincidence, but they would portray it as little more. Believing material structures inexorable, they would minimize the turn back from modernity and consider Charles V to be but the final paroxysm of a dying Middle Ages. Interests in sovereign statehood, in their view, were the culmination of centuries of material change. And most social scientific accounts of the state and the state system indeed claim the sufficiency of material change.

I am not the first to point out the influence of the Reformation upon Westphalia; some historians and a few social scientists have done so as well.[24] But none of them has described the mechanisms of this influence systematically or demonstrated its independence from material forces.

My own work, heeding the normal entanglement of ideas and material forces, argues for the Reformation not as a sole cause but as a central cause, by which I imply the following counterfactual claim: had the Reformation not occurred, a system of sovereign states would not have developed, at least not in the same form or in the same era as it did. More precisely, were it not for the Reformation, persistently medieval features of Europe—the substantive powers of the Holy Roman Empire and its emperor, the formidable temporal powers of the church, religious uniformity, truncations of the sovereign powers of secular rulers, Spain's control of the Netherlands—would not have disappeared when they did, to make way for the system of sovereign states.

[24] For historians, see Holborn (fn. 10); and Barraclough (fn. 10). For social scientists who acknowledge the influence of the Reformation and treat it briefly, see Bruce Porter, *War and the Rise of the State* (New York: Free Press, 1994), 68–69; and Hall (fn. 1), 51–67. Other scholars mention the role of other nonmaterial factors such as philosophical discourses, property rights, and the rise of Roman law. See Ruggie (fn. 1, 1993); and Bartelson (fn. 1).

To be valid, a counterfactual claim of this sort must follow a "minimal-rewrite-of-history rule," posing an alternative world where the alleged cause is extracted while other events and conditions remain intact. It must also meet a "cotenability" criterion, positing a theory of how, in this plausible alternative world, the remaining events and conditions might have brought about the same result.[25] Otherwise, the counterfactual is no more than an implausible straw man. It is not hard to envision an alternative world, one without the Reformation. It is precisely the one described by the dominant structural materialist accounts, a world where the state and the states system form completely apart from Protestantism.

But while such a world meets the standard of plausibility, it is not compelling. Without the Reformation, history would have looked different. In essence, medieval impediments to a sovereign state system would not have disappeared when they did but only much later. Nor would they have receded as sweepingly. Protestant ideas—continental in geography and encompassing in their content—challenged all temporal powers of the church and the empire. Sweeping causes produced sweeping effects. Were it not for the Reformation, the recession of the Middle Ages would likely have been staggered, this feature disappearing at one time, that feature disappearing decades or a century later. Material accounts themselves support this claim, for they show different states developing at different times.[26] With Hendrik Spruyt, I argue that "competitor" institutional forms persistently challenged the sovereign state long after the High Middle Ages, right up to the consolidation of the state. Spruyt places this consolidation roughly at Westphalia but does not treat the Reformation as an important explanatory cause.[27] My argument is that apart from the Reformation, institutional eclecticism would have continued long after Westphalia.

Had the Reformation not brought about the state system at Westphalia, much subsequent history would have been different, too. For out of the religious wars wrought by the Reformation came the hollowing of the Holy Roman Empire, the rise of the eighteenth-century balance of power, the Enlightenment, and the increased separation of

[25] On these criteria, see Philip E. Tetlock and Aaron Belkin, "Counterfactual Thought Experiments in World Politics," in Philip E. Tetlock and Aaron Belkin, eds., *Counterfactual Thought Experiments in World Politics* (Princeton: Princeton University Press, 1996), 21–25. With regard to the "minimal-rewrite-of-history rule," I will ultimately argue that even structural material forces were themselves shaped in part by the Reformation and were thus not independent of it. I claim here only that a world in which they were independent is plausibly imagined and indeed posited by most of the social scientists whom I address.

[26] See Tilly (fn. 7).

[27] Spruyt (fn. 1), 153–80.

religion and politics. Had these events not occurred, much else would have been different, too, and so on, with the trajectory of each historical branch becoming ever more speculative.[28] In reasoning counterfactually, we cannot be precise about this alternative, imaginary world. We can claim only that if a system of sovereign states had arrived within it, we would not call that system Westphalian but would instead associate it with some later century, some unimaginably different combination of wars and historical forces.

How can we demonstrate the dependence of Westphalia's timing and sweep upon the Reformation? The method here focuses on interests in Westphalia as they developed (or failed to develop) in separate polities, thus simplifying the complex causes of a continent-wide event and identifying an array of separable cases that can be compared across time and space.[29] Some of these polities—the German states and the northern provinces of the Netherlands—were protostates, partially independent: for them and for all those similarly situated, Westphalia meant independence. Others—France and Sweden—were already independent. Their interest took the form of sovereign statehood for the rest of Europe. All of these polities, as I show, developed their interest in sovereign statehood as a direct result of a crisis brought on by the Reformation. Using a positive strategy, I demonstrate the tight causal link between Reformation crises and the development of these interests. Then, negatively, I expose the weak connection between the same interests and structural material trends, which shape the skeptic's imaginary counterfactual world. If they fail to explain interests in Westphalia, then the imaginary world is not a compelling one. What the two strategies together support is a claim for what James March and Johan Olsen call a "historically inefficient" path to Westphalia, that is, an explanation of how norms and institutions develop in a way that cannot be predicted from prior environmental conditions, here, material ones.[30] Only an unexpected revolution in ideas can explain Westphalia.

[28] On counterfactuals, see James D. Fearon, "Counterfactuals and Hypothesis Testing in Political Science," *World Politics* 43 (January 1991); and Tetlock and Belkin (fn. 25).

[29] Gary King, Robert O. Keohane, and Sidney Verba refer to this strategy as "making many observations from the few"; see King, Keohane, and Verba, *Designing Social Inquiry: Scientific Inference in Qualitative Research* (Princeton: Princeton University Press, 1994), 217–28. Tetlock and Belkin (fn. 25) propose that this method can be used compatibly with and within counterfactuals (pp. 30–31). It meets what they call the "projectability" criterion.

[30] James G. March and Johan P. Olsen, "The Institutional Dynamics of International Political Orders," *International Organization* 52 (Autumn 1998), 958.

Two Roles of Ideas

Like scholarship on the Reformation, theories of the social influence of ideas are hardly new—Weber's essay on Protestantism is an example of both. Recently, though, scholars of international relations and comparative politics have led a revival of interest in ideas.[31] In explaining how the Reformation brought about Westphalia, I draw upon their work and that of earlier scholars in arguing that ideas exert power and mold political interests through two separate roles.[32]

In the first role, groups of people adopt new ideas into their identities. They undergo "conversion" and come to want what ideas prescribe. Identity here refers to a person's conception of which of his characteristics make him distinct from others according to his social role: is he a Lutheran, a Catholic, a German nationalist?[33] Identities are made up in part of ideas, which people hold stably over the long term. A person with a Protestant identity, for instance, persists in holding Protestant ideas. But identities can change and do so when people come to hold new ideas and self-conceptions.

That ideas form identities may seem like a trivial claim. Or even true by definition: do not ideas themselves compose identities? The claim becomes more meaningful, though, if we understand it as one about how people arrive at their new identities, that is, through reasoned reflection upon ideas' propositions about what is intrinsically valid, whether theoretically or practically. Such "reason of reflection" contrasts indeed with significant skeptical alternatives. It denies that people might adopt ideas as mere "hooks"—as the means to an entirely separate end, for example, wealth and position.[34] This approach disputes

[31] For general accounts of the trend, see Judith Goldstein and Robert O. Keohane, "Ideas and Foreign Policy: An Analytical Framework," in Goldstein and Keohane (fn. 3); Peter Katzenstein, ed., *The Culture of National Security: Norms and Identity in World Politics* (New York: Columbia University Press, 1996); Peter Hall, ed., *The Political Power of Economic Ideas* (Princeton: Princeton University Press, 1989); Wendt (fn. 1, 1994 and 1992); Mark Blyth, "Any More Bright Ideas? The Ideational Turn of Comparative Political Economy," *Comparative Politics* 29 (January 1997); John Kurt Jacobsen, "Much Ado about Ideas: The Cognitive Factor in Economic Policy," *World Politics* 47 (January 1995); Michael Mann, *The Sources of Social Power*, vol. 1 (Cambridge: Cambridge University Press, 1986).

[32] I propose my framework as a useful one for explaining the revolution at hand, not as the only process by which ideas operate. For portfolios of pathways and mechanisms, see Goldstein and Keohane (fn. 31), 8–26; and Katzenstein (fn. 31), 52–65. I am not proposing a general theory that denotes the conditions under which ideas will have influence; rather, the argument is a theory of the causes of revolutions in sovereignty (revolutions in ideas) and not one of what causes revolutions in ideas.

[33] For a definition of identity, see Ronald L. Jepperson, Alexander Wendt, and Peter J. Katzenstein, "Norms, Identity, and Culture in National Security," in Katzenstein (fn. 31): "the images of individuality and distinctiveness ('selfhood') held and projected by an actor and formed (and modified over time) through relations with significant 'others'" (p. 59).

[34] Kenneth Shepsle, "Comment," in R. Noll, ed., *Regulatory Policy and the Social Sciences* (Berkeley: University of California Press, 1985).

even a somewhat stronger role for ideas, that they "resolve uncertainty" about the means by which material ends are to be pursued but do not shape the content of ends themselves.[35] It also proposes identity change as more than a psychological mechanism for processing information.[36] Reason of reflection is identity change of the most fundamental sort.

My argument is not that reason of reflection is the only way in which people arrive at identities. The skeptical alternatives are plausible and doubtless actual; any major social revolution in identities quite probably involves some combination of the alternatives and reason of reflection. What I dispute is that identity change reduces to any of the skeptical alternatives. And I hypothesize that reason of reflection was a crucial motor behind the Protestant Reformation. Evidence for it lies in the correspondence of identity change with the rise of social circumstances that, by their very nature, promote reflection—the sermon, the pamphlet, the printing press, changes in theology, and the crisis in the Catholic church, among others. Evidence against a skeptical reduction lies in the lack of a tight link between such identity change and changes in the sort of social circumstances that foster the novel pursuit of separate, material ends, for example, class revolution.[37]

But shaping group identities is not enough. Popular conversion alone does not explain how ideas bring leaders of principalities or states to campaign or move troops on behalf of new international constitutions or how ideas hamper or overthrow obstructing leaders. In the second role of ideas, as social power, converts to new identities attempt to alter the costs and benefits to their prince, parliament, president, or anyone else who can deliver or hinder the new constitution. To this end they use carrots and sticks in the form of traditional political currency: votes, taxes, bureaucratic power, the threat to rebel, and so on. In this sense, the second role of ideas resembles rational choice theory, as people pursue their demands by altering the opportunity costs of those in a position to bring about their ends. What is distinctive here is that those wielding these tools are converts to new identities composed of ideas that define their preferences. Ideas here are manifestly more than mere legitimations for the pursuit of wealth or power and more, too, than "focal points" in bargaining.[38] The new social power is evidenced when

[35] Goldstein and Keohane (fn. 31), 16.

[36] See Brian Ripley, "Psychology, Foreign Policy, and International Relations Theory," *Political Psychology* 14 (September 1993).

[37] On processes of social construction, see Peter L. Berger and Thomas Luckmann, *The Social Construction of Reality* (New York: Doubleday, 1966); Finnemore and Sikkink (fn. 1), 913.

[38] Goldstein and Keohane (fn. 31), 17–20; Geoffrey Garrett and Barry R. Weingast, "Ideas, Interests, and Institutions: Constructing the European Community's Internal Market," in Goldstein and Keohane (fn. 3); Shepsle (fn. 34); for a key source on social power, see also Mann (fn. 31).

resultant outcomes (like interests in Westphalia) correspond in time and place to the strategic actions of bearers of identities asserting the demands implied by those identities.

Together, the two roles are the mechanism of contradiction between the new ideas and the old constitution, the fuel that spurs ideas from the heads of intellectuals to fresh international orders. The roles say little, though, about the people and the institutions involved. Ideas are not choosy about the "courier" that delivers them into politics; the roles are about processes, not actors. The cases at hand involve intellectual communities, transnational networks, ruling elites, and all classes of citizens who exercise their power by changing their pattern of religious participation, joining Protestant armies, threatening rebellion, and otherwise creating incentives for leaders to pursue Westphalian sovereign statehood.[39] These actors and forms of power express new kinds of interest differently in different polities through (1) "Reformation from below," in which social groups pressured elites; (2) "Reformation from above," in which elites played a stronger, more proactive role; and (3) the "politique solution," in which Protestant reformers did not succeed in creating a Protestant state but coaxed a social compromise that entailed the political secularization embodied in Westphalia.[40]

Each of the two roles of ideas—as changing identities and as wielding social power—echoes the commitments of constructivist scholars. In explaining how interests are shaped, some constructivists embrace the first role of ideas as transforming identities. Far more, though, stress the second role, of how actors or institutions whose identities are defined by ideology and culture then exert social power to shape the state's interests.[41] The framework employed here acknowledges and combines the two strands. Like the constructivists, this account also takes issue with those rationalist theories—neorealism and neoliberal-

[39] On intellectual communities, see Emanuel Adler and Peter Haas, "Conclusion: Epistemic Communities, World Order, and the Creation of a Reflective Research Program," *International Organization* 46 (Winter 1992). On transnational networks, see Kathryn Sikkink, "Human Rights, Principled ssue Networks, and Sovereignty in Latin America," *International Organization* 47 (Summer 1993); nd Margaret Keck and Kathryn Sikkink, *Activists beyond Borders: Advocacy Networks in International Politics* (Ithaca, N.Y.: Cornell University Press, 1998). For an example of the influence of publics, see David Lumsdaine, *Moral Vision in International Politics* (Princeton: Princeton University Press, 1993).

[40] On causal pathways, see Goldstein and Keohane (fn. 31), 24–26; and Katzenstein (fn. 31), 52–65.

[41] In the first category, see Mlada Bukovansky, "American Identity and Neutral Rights from Independence to the War of 1812," *International Organization* 51 (Summer 1997); Richard Price, "Reversing the Gun Sights: Transnational Civil Society Targets Land Mines," *International Organization* 52 (Summer 1998); Richard Price and Nina Tannenwald, "Norms and Deterrence: The Nuclear and Chemical Weapons Taboo," in Katzenstein (fn. 31); Audie Klotz, *Norms in International Relations: The Struggle against Apartheid* (Ithaca, N.Y.: Cornell University Press, 1995); Finnemore and Sikkink (fn. 898. In the second category, see, for examples, the essays in Goldstein and Keohane (fn. 3); and most of the essays in Katzenstein (fn. 31).

ism—that assume that states have fixed identities and interests, usually material, and that they pursue these interests through rational calculations of ends and means. Rather, states' identities and interests are shaped through their social environment.[42] But we should not draw too sharply the contrast with rationalist accounts, the more sophisticated versions of which do take account of the role of identities, norms, and social contexts.[43] It is only with those rationalist theories that take a structural materialist form that constructivism and the account presented here part ways. Most accounts of the formation of the state and the states system indeed express such a materialist rationalism.

In one respect, though, I propose to extend the constructivist agenda by explaining international constitutive normative structures (here, the structure of sovereign constitutional authority) as the result of the actions of agents (here, states). Although constructivists acknowledge the importance of sovereignty as a constitutive norm, they typically treat it as a determinant of state identities and behavior, and not as their product.[44] There are exceptions to this thinking, however, most prominently Rodney Bruce Hall's *National Collective Identity*, a work that traces the relationship between national identity and the international system from early modern Europe through the First World War. Broadly sympathetic to Hall, my argument focuses on what he calls "territorial sovereignty" and its origins in the Protestant Reformation, an event that he treats far more briefly than I do.[45] Friendly commentators on constructivism call for just the sort of theoretical development that authors like Hall have begun and that I seek to further: Jeffrey Checkel urges an explanation of how agents shape normative structures; Janice Thomson commends an account of the origins of norms of sovereignty.[46]

[42] See Jepperson, Wendt, and Katzenstein (fn. 33), 58–60.

[43] Such versions adopt a variety of strategies, ranging from showing how utility functions are specified by ideas, culture, or psychological schema, to asserting the rationality of the attempts of "norms entrepreneurs" to construct common knowledge and alter others' utility functions in accordance with their commitments, to devising models of how ideas modify the pursuit of rational action as "focal points" or "resolvers of uncertainty," and to charting the social context of rational action. See Finnemore and Sikkink (fn. 1), 909–15; Goldstein and Keohane (fn. 31), 3–30; Miles Kahler, "Rationality in International Relations," *International Organization* 52 (Autumn 1998), 933–38; March and Olsen (fn. 30), 952–54; Jon Elster, *Nuts and Bolts for the Social Sciences* (Cambridge: Cambridge University Press, 1989); idem, *Political Psychology* (Cambridge: Cambridge University Press, 1993). For a more skeptical view of the reconcilability of constructivist and rationalist traditions, see Ruggie (fn. 1, 1998), 883–85.

[44] See the essays in Biersteker and Weber (fn. 1); Jepperson, Wendt, and Katzenstein (fn. 33), 45–46; Ruggie (fn. 1, 1993); Ruggie (fn. 1, 1998), 870; Wendt (fn. 1, 1992), 412–15.

[45] Hall (fn. 1), 51–58.

[46] Checkel (fn. 1), 340–42; Thomson (fn. 1). See also John Gerard Ruggie's comment that "[s]ocial constructivists in international relations have not yet managed to devise a theory of constitutive rules." Ruggie (fn. 1, 1998), 872.

Structural Skepticism of the Reformation's Influence

Wary of these alleged workings of ideas are critics who would explain polities' interests by appealing instead to various material structures. These are those social constraints on the behavior of individuals and groups that do not consist of ideas themselves—technology, institutions of coercion and taxation, class, societal coalitions, the international distribution of military and economic power, and the like. According to these critics, polities developed an interest in a sovereign state system when they accrued the power inhering in arms, wealth, stronger extractive bureaucracies, and the like. Interests, that is, followed from power and, in turn, enabled polities to prevail over rival institutions like the Holy Roman Empire. Westphalia was merely the set of norms that legitimated their ultimate victory.

According to some versions of this structuralism, the state prevailed because it was the institution best suited to fight wars. Examples are the work of Charles Tilly and of the theorists of the "military revolution," the early modern transformation of military technology and organization that rendered states the most optimal institution for bearing the cost and scale of warfare.[47] Other versions see the state triumphing because it could best produce wealth, through its ability to assure property rights, extract taxes, and commit credibly to trade agreements with other states. Hendrik Spruyt, for instance, argues that an explosion of trade during the Middle Ages shifted social power toward the towns, which then allied with monarchs to create state institutions that could best foster commerce.[48] In still other versions, the state developed out of changes in the early modern class structure.[49]

Finally, there are realist accounts of the triumph of the state. This, at first, may seem misleading: does not realism assume that sovereign

[47] Otto Hintze, *The Historical Essays of Otto Hintze* (Oxford: Oxford University Press, 1975); Richard Bean, "War and the Birth of the Nation-State" *Journal of Economic History* 33 (March 1973); Porter (fn. 24); Tilly (fn. 7). The exception is Hall (fn. 1), who treats the Reformation as an important source of territorial sovereignty. Porter is also a qualified exception here, for he acknowledges the partial explanatory role of the Reformation. See Porter (fn. 24), 68–69. The locus classicus on the military revolution is Roberts (fn. 16, 1967), 195–225. An impressive recent work on the long-term political results of the military revolution is Brian Downing, *The Military Revolution and Political Change* (Princeton: Princeton University Press, 1992), esp. 56–74. For a critique of the military revolution thesis, see Geoffrey Parker, "The 'Military Revolution,' 1550–1660: A Myth?" *Journal of Modern History* 48 (June 1976); and idem, *The Military Revolution* (Cambridge: Cambridge University Press, 1988).

[48] Douglass C. North and R. P. Thomas, *The Rise of the Western World* (Cambridge: Cambridge University Press, 1973); Spruyt (fn. 1).

[49] Immanuel Wallerstein, *The Modern World System: Capitalist Agriculture and the Origins of the European World-Economy in the Sixteenth Century* (New York: Academic Press, 1974); N. K. Anderson, "The Reformation in Scandinavia and the Baltic," in G. R. Elton, ed., *The New Cambridge Modern History* (Cambridge: Cambridge University Press, 1958).

states already exist? Strictly, yes, but we can still find accounts closely resembling realism that tell how the European balance of power shifted away from the empire, Spain, and the Habsburgs and toward an alliance of France, Sweden, and the partially independent German and Dutch protostates. Westphalia merely ratified the victory of this alliance.[50] In sum, any account rooted in ideas must answer all of these forms of skepticism.

EVIDENCE FOR THE REFORMATION AS ORIGIN OF WESTPHALIA

Reformation crisis as used here means a deep and contested social struggle between advocates of a Protestant political order and advocates of a Catholic political order within a polity, sometimes involving widespread violence but always involving a demand on both sides for uniformity of faith. It was through such a Reformation crisis that polities came to develop an interest in sovereign statehood. The social struggle is what unfolded through the two roles of ideas. But to know how Protestant ideas wielded influence, we must first know what Protestant ideas were and why they were politically potent. The case for the influence of the Reformation begins with the answers to these questions.

THE POLITICAL THEOLOGY OF THE REFORMATION

Though voluminous and complex and rendered variously, Reformation theology uniformly asserted a common doctrine about what the Christian church was not: a visibly united institution under a single human authority. At least for Martin Luther, it was simply the aggregate of doctrinally authentic local churches.[51] This ecclesial theology had enormous institutional implications—confiscating the church's large tracts

[50] See Paul Kennedy, *The Rise and Fall of the Great Powers: Economic Change and Military Conflict from 1500 to 2000* (New York: Random House, 1987), 31–73; Aristide R. Zolberg, "Origins of the Modern World System: A Missing Link," *World Politics* 33 (January 1981), 253–58; Ludwig Dehio, *The Precarious Balance*, trans. Charles Fulman (New York: Knopf, 1962); S. H. Steinberg, "The Thirty Years War and the Conflict for European Hegemony," in Rabb (fn. 13); Robert Gilpin, *War and Change in World Politics* (Cambridge: Cambridge University Press, 1981).

[51] Martin Luther, *The Freedom of a Christian*, trans. W. A. Lambert and ed. H. J. Grimm, in *Luther's Works*, vol. 31 (Philadelphia: Westminster Press, 1957); idem, *Bondage of the Will*, trans. Philip S. Watson and Benjamin Drewery and ed. P. S. Watson, in *Luther's Works*, vol. 33 (Philadelphia: Westminster Press, 1972); Quentin Skinner, *The Foundations of Modern Political Thought*, vol. 2 (Cambridge: Cambridge University Press, 1978); Euan Cameron, *The European Reformation* (Oxford: Oxford University Press, 1991), 99–198; Heiko Oberman, *Luther: Man between God and the Devil* (New Haven: Yale University Press, 1989); Steven Ozment, *Protestants: The Birth of a Revolution* (New York: Doubleday, 1992), 43–86; Louis Bouyer, *The Spirit and Forms of Protestantism* (London: Harville Press, 1955).

of property, ending the emperor's authority to enforce religious uniformity, forbidding ecclesiasts from holding temporal offices, and eliminating the ecclesiastical powers of the pope and emperor.[52]

But such relinquished powers could not be left adrift. They were assumed by the secular authorities—in Germany, by princes; in the Netherlands, by the Estates-General; in Sweden and England, by the king. This new separation of functions also sprang from Luther's "Doctrine of the Two Kingdoms and the Two Governments," his political theology. God created two earthly orders with two forms of government. One was the realm of the spirit, the site of the relationship between Christ and the believer's soul; the other was the realm of the world, the order of secular society, governed through civil magistrates, laws, and coercion. The reformers demanded a separation. Thus, the pastors of the church were not to perform the duties of public order, just as magistrates, princes, and kings would not preach or perform the sacraments.[53] In separating the two realms, Reformation political theology essentially prescribed sovereignty, even though neither Luther's tracts nor John Calvin's *Institutes* outline a Westphalian system of sovereign states. For secular authorities within the empire the remaining temporal prerogatives of the church filled out their portfolios of power. The point is essential: sovereignty was implied in the very propositions of the Reformation.

The church condemned Luther at the Diet of Worms of 1521; a decade later Charles V sent his troops to quell the heresy, an imperial endeavor that would continue at least through the 1630s. Facing armed threat, the reformers now found even more reason to ally with and give full sovereignty to the secular rulers, whose armies could protect them. Reformation theology, then, led to sovereignty not only directly, by prescribing it, but also indirectly, by motivating Charles V to quash its challenge and by prompting reformers, in turn, to seek protection. Although physical protection considered in isolation is hardly a theologically laden desideratum, the reformers' need for it is scarcely intelligible apart from their heretical beliefs, which alone had elicited the emperor's armed response. Sovereign statehood was the carapace that would stanch the Counter-Reformation.

[52] Cameron (fn. 51), 145–55.

[53] Ibid., 152–53; Martin Luther, *Whether Soldiers, Too, Can Be Saved*, trans. Charles M. Jacobs and ed. Robert C. Schultz, in *Luther's Works*, vol. 46 (Philadelphia: Westminster Press, 1967); idem, *Temporal Authority: To What Extent It Should Be Obeyed*, trans. J.J. Schindel and ed. W. I. Brandt, in *Luther's Works*, vol. 45 (Philadelphia: Westminster Press, 1967); Ozment (fn. 51), 122–40; Skinner (fn. 51), Romans 13.

Another important feature of the Reformation was the novelty of its political propositions in the history of European Christendom. Such originality parries a methodological objection: if it was heretical theological ideas that led to Westphalia, then why did previous heresies not have the same political effect? Indeed, to the contrary, heresies of the preceding three centuries—the Lollard movement in England, the Waldenses in southern Europe, and the Hussites in Bohemia—did not imply the same political ideas. The earlier movements challenged church doctrines and practices, to be sure, but did not advance a political theology that undermined the church's hierarchical or temporal powers.[54] True to the argument, different ideas with a different content produced a different political result.

The Correlation between the Reformation and Polities' Interests in Westphalia

If Reformation theology and sovereignty are connected intrinsically and conceptually, they are also connected historically and causally. Thus, every polity that came to have an interest in a system of sovereign states had experienced a strong Reformation crisis, whereas in every polity that fought against a sovereign states system, the Reformation won few converts. The correlation also involves timing: most of the polities that acquired an interest in a system of sovereign states usually did so within a generation of the arrival of Protestantism in their land; most of then had shown little inclination toward a system of sovereign states befor Protestantism arrived. These correlations anchor the argument.

The first three columns of Table 1 show this correlation simplified Column 1 lists the major polities in Europe during the century prior t Westphalia. Column 2 displays whether the polity adopted an interes in a system of sovereign states and if so, when. Column 3 then display whether and when a polity experienced a Reformation crisis. Column 4 generalizes about the timing and nature of the growth in the state's institutions. Columns 2 and 3 show that the presence and timing of a Reformation crisis correspond strongly with a political interest ir Westphalia. In the four chief polities (or region of polities, in the cas of Germany) that fought for Westphalia in the Thirty Years' War—Germany, the Netherlands, Sweden, and France—an interest in a system of sovereign states arose within a generation of its Reformatior crisis. Though not integrally involved in the armed conflict for Westphalia, England, Denmark, and Transylvania, too, experienced a Refor-

[54] Cameron (fn. 51), 70–78.

TABLE 1

THE TIMING OF INTERESTS IN WESTPHALIA AND THEIR ALLEGED CAUSES

(1) Polity	*(2) Interest in Westphalian System of Sovereign States*	*(3) Reformation Crisis*	*(4) Growth in Material Power*
German Protestant states[a]	1520–45	1520–1540s	gradual in 16th century; sharp after 1650
German Catholic states[a]	none	weak, remaining minority status	gradual in 16th century; sharp after 1650
Netherlands (United Provinces)	1581	1560s	sharp beginning in 1590s
Sweden	1540s	1520s	gradual after 1600; sharp between 1672 and 1718
France	1620s	1560s–1598	gradual and halting after 1600; sharp after 1650
Spain	none	none	sharp from 1470s to 1550s
Italy	none	none	1400s, growth of city-states, approximation of sovereign states system until 1527
Poland	none	spread of Protestantism but no crisis	little significant growth
Habsburg Hungary	none	mild spread of Protestantism but no significant crisis	little significant growth
Transylvania	1618	1550	little significant growth
Denmark	1538	1520s and 1530s	little significant growth
England	support for European Protestants short of all-out war, 1560s–1648	1530s	significant growth through 16th century

[a]Some German states switched between official Catholic and Protestant status as a result of war and political bargaining between the 1555 Peace of Augsburg and 1648. I label the states here according to the status they first assumed during the Reformation crisis of 1520–45.

mation crisis and supported the anti-imperial powers diplomatically. Important, too, is the fact that none of the Catholic polities, the Catholic German principalities, Spain, Italy, or Poland, developed any interest at all in a system of sovereign states. They remained allies with the empire.

Princes and kings came to pursue an interest in a system of sovereign states when and where a Reformation crisis took place. This is important evidence, but circumstantial. Who, then, were the players, what were the events and institutions, through which the two roles of ideas advanced?

Conversion to Protestantism, the first role of ideas, is evidenced in the new religious practices of those who adopted the Reformation. Protestants attended churches where preachers taught Protestant theology; they conducted mass, confession, and the other sacraments as the new teaching prescribed; the more unruly of them smashed icons and rioted against Catholic churches; they refused to obey the sanctioning authority of Rome. But what brought them to conversion? It was the theology of Martin Luther that led most immediately to the Reformation crisis. Luther did not create his ideas de novo but was influenced by late medieval theology, especially nominalism. His own spiritual crisis was an important precipitant, too, as his ideas circulated quickly and alarmingly first through the monasteries, then through the clergy and missionaries, then to congregations—a pattern of diffusion replicated throughout Europe, in the many versions that Protestant ideas assumed.[55]

Two main conduits, both of them fairly novel, aided this dissemination—the sermon and the pamphlet. Protestant theology and worship placed new importance on the pastor's conveyance of the Word, an imprimatur of authority that attracted its hearers to its injunctions.[56] In the same period the print media burgeoned. Between 1517 and 1520 Luther's followers printed thirty of his writings in well over three hundred thousand copies, pamphlets devoured in the universities and by clerics, who then spread their ideas through word of mouth to larger audiences. Pamphlets proliferated in other locales as well.[57] Both modes of communication, sermon and pamphlet, were received in virtually all

[55] Ibid., 99–110.
[56] Robert Wuthnow, *Communities of Discourse* (Cambridge: Harvard University Press, 1989), 131.
[57] Ibid., 129–40. On the importance of print media, see Elizabeth Eisenstein, "The Advent of Printing and the Protestant Revolt: A New Approach to the Disruption of Western Christendom," *Annales, E.S.C.* 26 (1971); Richard G. Cole, "Propaganda as a Source of Reformation History," *Lutheran Quarterly* 22 (1970); C. S. L. Davies, *Peace, Print, and Protestantism* (London: Paladin, 1977).

sectors of society, but most enthusiastically by merchants and artisans, who made up the majority of Protestants.

Other social precipitants abetted the Reformation, too. Most obvious and familiar was the perceived corruption of the medieval church. In Germany, where the Reformation began, this perception was stronger than anywhere else in Europe and had been growing steadily over the decades prior to the Reformation.[58] Intellectual, theological movements also corresponded to the Reformation. Christian humanism and nominalism, both precursors of Reformation theology, were most widespread in Germany, Switzerland, and the Low Countries, where the Reformation happened earliest and most intensely, and had been steadily building over the previous two centuries.[59] The influence of other factors—the printing press, literacy, the intensity of central papal and imperial control—is witnessed in urban settings, where the Reformation was most successful.[60]

Reformation scholars also emphasize a separate urban factor, the economic interests of the rising bourgeoisie. The debate over the relationship between Protestantism and capitalism, of course, dates back to Friedrich Engels and Max Weber; it is not one that I intend to join here. Few Reformation historians today, though (even among the recent generation that emphasizes the social causes of Protestantism), argue that Protestant ideas were reducible to, rather than the product of, a complex interaction with class and economic interests. Few would argue away the autonomy of Reformation ideas.[61]

Once converted, Protestants then brandished their social power—the second role of ideas—in diverse ways. Since the church's authority depended most directly on the believer's practice, its authority was most

[58] Steven Ozment, *The Age of Reform, 1220–1550: An Intellectual and Religious History of Late Medieval and Reformation Europe* (New Haven: Yale University Press, 1980), 222–23; Ozment (fn. 51), 9–86; Cameron (fn. 51), 20–37, 79–93.

[59] Alistair McGrath, *The Intellectual Origins of the European Reformation* (Oxford: Basil Blackwell, 1987); idem, *Reformation Thought: An Introduction* (Oxford: Oxford University Press, 1988); Heiko Oberman, *The Harvest of Medieval Theology: Gabriel Biel and Late Medieval Nominalism* (Cambridge: Harvard University Press, 1963); idem, *Masters of the Reformation: The Emergence of a New Intellectual Climate in Europe*, trans. Dennis Martin (New York: Cambridge University Press, 1981).

[60] Wuthnow (fn. 56), 25–51; Andrew Pettegree, *The Early Reformation in Europe* (Cambridge: Cambridge University Press, 1991); Bernd Moeller, *Imperial Cities and the Reformation*, trans. H. C. Erik Middlefort and Mark U. Edwards, Jr. (Philadelphia: Fortress Press, 1972); Thomas Brady, "In Search of the Godly City: The Domestication of Religion in the German Urban Reformation," in R. P.-C. Hsia, ed., *The German People and the Reformation* (Ithaca, N.Y.: Cornell University Press, 1988).

[61] On Reformation historiography, see Steven E. Ozment, ed., *Reformation Europe: A Guide to Research* (St. Louis, Mo.: Center for Reformation Research, 1982); Hsia (fn. 60); Wuthnow (fn. 56), 25–51; Pettegree (fn. 60); and Bob Scribner, Roy Porter, and Mikulcs Teich, eds., *The Reformation in National Context* (Cambridge: Cambridge University Press, 1994).

directly undermined when believers simply practiced heterodoxy. Protestants attended churches where preachers taught Protestant theology; they conducted mass, confession, and the other sacraments as the new teaching prescribed; the more unruly of them smashed icons and rioted against Catholic churches; and they refused to obey the sanctioning authority of Rome.[62] Defiant practice was itself a form of social power.

Practical dissent ultimately required a coarser power, too. Protestants took up arms against the emperor on behalf of the prince who would ally with them, and they rebelled against the resistant prince. Merchants, laborers, artisans, peasants, and nobles of all ranks offered their military service in one of these ways. In every case any form of social power had to provide the prince, king, or magistrate with the incentive and means to seize the church's lands and temporal powers and perhaps also to fight the emperor. If the head of a polity himself converted, then he could exercise the social power of ideas by virtue of his position.[63] He could enforce doctrinal uniformity within his territory, appoint new church leaders, nullify the authority of the Roman hierarchy, disband the monasteries, seize church property, and raise troops to fight the empire. But even if he did not convert, out of either obstinacy or indifference, he might still further the Reformation in these ways, either because he feared the rebellion of converts or because he saw an opportunity to gain wealth and power through allying with them. Whether he was pious, rapacious, opportunistic, or afraid, it was only because of the extensive Protestant social power arrayed around him that a leader could successfully defy the church.[64] The social power of position wielded by heads of polities, along with the social power of defiant religious practice, of rebellion, and of military service, together constituted the social power of the Reformation.

This is a general account of how Reformation ideas, through their two roles, brought polities to their interest in sovereign statehood. Different polities, though, experienced variations on this narrative. They followed three causal pathways, each of which helps us to see more precisely how and through whom conversion and social power operated: Reformation from below, Reformation from above, and the politique solution. Figure 1 presents a schematic summary of each of the path-

[62] Cameron (fn. 51), 99–110, 199–318; Ozment (fn. 51), 43–86.

[63] For an example of the influence of a ruling elite, in this case Gorbachev, who himself converts to and empowers new ideas, see Robert G. Herman, "Identity, Norms, and National Security: The Soviet Foreign Policy Revolution and the End of the Cold War," in Katzenstein (fn. 31).

[64] Cameron (fn. 51), 199–318.

1. Reformation from Below (Germany, the Netherlands, Transylvania)

Pathway of Conversion

Reflection of theologians → monasteries and universities → ordinary clergy → conversions in a broad array of sectors, including ruling elites

Pathway of Social Power

Broad array of Protestant social sectors especially urban middle classes, exercising social power of defiant religious practice, resistance, and participation in armed forces (social crisis) → elites → the polity's interest in Westphalia

2. Reformation from Above (Sweden, England, Denmark)

Pathway of Conversion

Reflection of theologians → monasteries and universities → missionaries → ordinary clergy

↓

conversions in a broad array of sectors, including ruling elites

Pathway of Social Power

Protestant ruling elites exercising the social power of their position (social crisis) → the polity's interest in Westphalia

↕

broad array of Protestant social sectors, empowering and empowered by ruling elites (social crisis)

3. The Politique Solution (France)

Pathway of Conversion

Reflection of theologians → monasteries and universities → missionaries → ordinary clergy

↓

conversions in a broad array of sectors, including nobility

Pathway of Social Power

Protestants in broad array of sectors exercising social power of armed resistance → interest in secular politique solution among powerful Protestant and Catholic elites → politique compromise whose foreign policy power includes an interest in Westphalia

FIGURE 1
SCHEMATIC SUMMARY OF EACH PATHWAY

ways, showing the chain of events through which the two roles of ideas unfolded. Reformation from below and the politique solution are then illustrated through the cases of Germany and France, respectively. Reformation from above and the experience of other European polities listed in Table 1 are described more briefly.[65]

GERMANY: REFORMATION FROM BELOW

In Germany, the Reformation began; in Germany, it first had political effects. Between 1523 and 1546 princes and magistrates scattered across more than half of the German-speaking territory, including Swiss and Austrian regions, secularized and seized church lands, cutting off the local church hierarchy from the commands of Rome and effectively creating a territorial church. Prepared to fight for what they seized, they resolved to defend an interest in sovereignty. As Figure 1 shows, ruling elites are the final link in the causal chain that ends with this interest; they are the ones whose position conferred upon them the social power to take it up. Rulers came to such interests only amidst the widespread conversion to Lutheran ideas, as converts were practicing religious defiance, threatening rebellion, and joining armed forces against the empire.[66]

Within a decade of Luther's public dissent in 1517, German rulers began to wrest extant prerogatives from imperial powers. These rulers, from three hundred separate polities in Germany, were an eclectic group of city leaders, nobles, knights, princes, and sometimes imperial electors. In Luther's own region of Saxony, for instance, the elector John placed church lands under state control, replaced Catholic bishops with a "visitation" committee of electoral councillors and theologians, and began to enforce devotional uniformity, improve the standards and conduct of the clergy, and oversee the moral and spiritual affairs of the laity.[67] Much the same occurred in dozens of cities and principalities across Germany between 1523 and the late 1540s, by which time the German Lutheran Reformation finally ceased to expand. The princes often took greater control over the church than

[65] The method here corresponds to the technique of "process tracing" in Alexander George's method of structured focused comparison; see George, "Case Studies and Theory Development: The Method of Structured, Focused Comparison," in Paul G. Lauren, ed., *Diplomacy: New Approaches in History, Theory, and Policy* (New York: Free Press, 1979).

[66] Holborn (fn. 10), 37–51, 137–39, 158, 162, 374; Owen Chadwick, *The Reformation* (London: Penguin, 1964); John M. Todd, *Reformation* (New York: Doubleday, 1971), 230–39; A. G. Dickens, *Reformation and Society in Sixteenth Century Europe* (London: Thames and Hudson, 1966), 87–106; G. R. Elton, *Reformation Europe, 1517–1559* (London: Collins, 1963); Barraclough (fn. 10), 262–67.

[67] Elton (fn. 66), 56; Chadwick (fn. 66), 67–71; Todd (fn. 66).

Luther had intended, although he was generally supportive of their appropriation. He actively assisted Elector John of Saxony in placing church governance in secular hands. Where princes made Protestantism official, they generally also made it uniform, requiring worship in a single church, outlawing dissent, and making Protestantism constitutive of their order's legitimacy. During the rest of the sixteenth century the reliance of Protestant churches on the prince only grew stronger.[68]

How did prince and town leaders come to wrest such power, create such an order? Only in response to what came first from below. Conversion in Germany was paradigmatic, flowing from theologians to monastery to university to clergy to congregation, involving the sermon, the pamphlet, and all of the circumstances that generally obtained in conversion across Europe, as described above and illustrated in Figure 1. The figure also illustrates the exercise of social power. In virtually every German city or principality that became Protestant, the Reformation first spread popularly, to merchants, laborers, shopkeepers, farmers, and sometimes nobles; only afterward did town councils or princes ally with reformers. Wherever Protestantism spread, it brought defiant religious practice almost immediately: priests and parishioners adopted new rites and rituals, ignored the commands and excommunications of their superiors in the Roman Catholic hierarchy, and often rioted and staged public protests. It was a popular movement that the ruler could ally with to his benefit or ignore at a cost.[69]

Converts also exercised more direct political power, according to patterns that varied from region to region. In many northern cities the most important converts were merchants and craftsmen who were excluded from political privilege. Once they converted, they attempted coups, exerting the social power of rebellion, often winning over acquiescent city leaders. The Reformation also reached the rural proletariat, which rose up in the peasant rebellions of the mid-1520s but was cruelly and summarily defeated in battle by the nobility, indeed with Luther's encouragement. In other cities, in central and southern Germany and northern Switzerland, a "civic reformation" took place in which conversion occurred across several layers of society, but most crucially among literate and influential economic elites. Through city politics, these elites succeeded in overcoming the opposition of other elites in persuading city leaders to support the Reformation. In almost every

[68] Barraclough (fn. 10), 374.
[69] Cameron (fn. 51), 199–318.

German principality that became Protestant and supportive of Westphalia, monarchs converted only after vast swaths of their publics had already done so themselves. In every Protestant polity, too, once the Reformation became official, converts exercised social power by joining armed forces to fight the empire. By contrast, in those German polities such as Bavaria that remained officially Catholic and allied with the empire throughout the sixteenth century, the Reformation spread comparatively little among the populace, was not taken up by heads of polities, and was fought by Catholics who correspondingly populated Catholic armies.[70]

Once polities began to become officially Protestant, Emperor Charles V and the stalwart Catholic princes were forced to respond. After several failed attempts at reconciliation during the 1520s, Catholic forces prepared to take up arms against Protestantism. The Protestant principalities organized into the Schmalkaldic League for defense, and Charles V determined to end the schism. From then on both sides would fight and a unified empire was doomed. In these struggles Protestant converts exercised social power by joining armies and fighting for their polity once it had become officially reformed. Of course, Catholics exercised similar social power by fighting for their ruler—those princes and magistrates who did not go along with the Reformation. Hesse, Brandenburg, Electoral Palatinate, Saxony, and tens of other principalities and cities became officially Protestant; rulers in Bavaria, Austria, and elsewhere remained Catholic; some regions switched back and forth during the sixteenth century. The 1555 Peace of Augsburg proved only a temporary truce, not a lasting solution. "Since conversions and reconversions continued to occur after 1555, Germany's confessional map resembled a periodically changing checkerboard into the time of the Thirty Years' War," historian John Gagliardo writes. Only after Westphalia would the sovereignty of German states rest unchallenged.[71]

In German polities, then, the Reformation came from below, with conversions across societal sectors and social power exercised through defiant religious practices, threats of rebellion, politics within cities and principalities, rulers asserting the power of their position, and broad participation in the armed forces. The timing of interests in sovereign

[70] Ibid., 210–92; on threat theory, see Stephen Walt, *The Origin of Alliances* (Ithaca, N.Y.: Cornell University Press, 1987).

[71] Holborn (fn. 10), 137–39, 158, 162, 284–95, 374; Chadwick (fn. 66), 67–71; Todd (fn. 67), 230–39; Dickens (fn. 66), 87–106; Elton (fn. 66); Barraclough (fn. 10); Gagliardo (fn. 10), 14; Cameron (fn. 51), 210–91.

statehood testifies to these influences. German assertions of sovereignty during the 1520s, 1530s, and 1540s corresponded remarkably well with the furious advance of the German Lutheran Reformation during this same period. Prior to Luther's defiance of 1517, German rulers had not threatened or hinted at a takeover of the church's temporal powers, though they had long appealed to German nationalism and brought hostile grievances against the church and empire before imperial diets. Not until after Luther rebelled did the first German ruler take up the cause and then not until his own subjects had first taken up reformed teachings. In Germany the coincidence in timing between the Lutheran Reformation and interests in sovereign statehood was virtually exact.[72]

Geographic comparisons also corroborate the same relationship between confessional loyalty and political interests. With very few exceptions, polities that were officially Protestant fought for sovereign statehood and independence from the empire, whereas officially Catholic polities remained loyal to the empire. In two periods of German war, between the 1530s and the 1550s and then from the 1580s through Westphalia in 1648, alliances of Protestant polities fought alliances of Catholic polities and the empire. Alliance patterns consistently followed confessional lines, even when a polity changed religious hands; typically, it would then also change its alignment. Apart from the spread of Protestantism, there appears to be no geographical pattern as to which polities became Catholic and which became Protestant. There is no pattern, for instance, that corroborates a proximity-of-threat hypothesis drawn from international relations scholarship, one that would predict lands on the border of Habsburg Austria to remain Catholic in order to attain safety from the Holy Roman Empire, and more distant, buffered lands to become Protestant. In fact, both Protestant and Catholic lands bordered Austria and were scattered throughout Germany without apparent geostrategic pattern.[73]

Against this case for ideas, the most immediately obvious skepticism claims instrumentality—that secular rulers used the Reformation to legitimate their seizure of sovereignty and their acquisition of the church's remaining powers and wealth. According to this line of thinking, the Reformation exercised little social power over rulers. For four reasons, though, such motives are not a persuasive substitute for ideas. First, Reformation historians reveal the rapacity of some princes and

[72] Holborn (fn. 10), 37–51; Barraclough (fn. 10), 363–67; Gagliardo (fn. 10), 2–4.
[73] See Cameron (fn. 51), 199–313.

rulers but also assert the veritable devotion of others. Elector Frederick the Wise, for example, supported and harbored Luther shortly after his famous defiance of the pope and emperor at Worms, a time when Luther had few other powerful public supporters.[74]

But even if one were to assume that every prince was motivated by familiar material desiderata, the argument encounters a second objection: it is unclear that ecclesial booty was a rational desideratum at all. Several historians of the Reformation agree that most reforming princes failed to acquire great wealth from church lands. Many, such as Frederick the Wise, risked their power, even their lives, for their stand. That the benefits outweighed the costs was hardly clear.[75] But even assuming that all rulers who sided with the Reformation did so for instrumentally rational, material reasons, there is a third problem: the explanation does not account for the sorts of social power that enabled princes to seize the church's powers and wealth. Who fought in their armies? Who occupied the churches and disobeyed the bishops? Rulers' motives alone fail to explain this. One needs to take account of the sources of their popular support.

The fourth—and most damaging—problem with the argument is, again, one of timing. It fails to explain why, if secular powers were motivated by what they stood to gain materially from church goods, they chose to seize sovereignty at this particular time and only in certain places. Why did some European princes and nobles rebel against the emperor, while others, similarly rich and empowered, remained loyal? And why did rebellious rulers not pursue sovereignty one hundred years earlier? Late medieval theology, after all, abounded with ideas that challenged traditional authority and that might well have served as legitimating rationales.[76] The argument from motives alone has little capacity to explain the historical specificity—the timing or the pattern—of rulers' interests in Westphalia. As I have argued, the presence or absence of Reformation ideas explains these interests quite well. The two roles of ideas tell the story behind these interests. In Germany it was a story of reformation from below.

FRANCE: THE POLITIQUE SOLUTION

France defined its interest in a European system of sovereign states quite differently than Germany did. As it was already a sovereign state,

[74] Ibid., 101–3.

[75] Ibid., 294–99; F. L. Carsten, *Princes and Parliaments in Germany from the Fifteenth to the Eighteenth Century* (Oxford: Oxford University Press, 1959); A. G. Dickens, *The German Nation and Martin Luther* (London: Edward Arnold, 1974); Brady (fn. 60).

[76] McGrath (fn. 59); Oberman (fn. 59).

it came to fight for sovereign statehood for the rest of Europe, entering the Thirty Years' War in 1635 to bring the defeat of the Holy Roman Empire's remaining sovereign powers in 1648. France nonetheless retained its Catholic monarchy. The lesson: a state need not become officially Protestant in order to pursue an interest in sovereign statehood; rather, it need only experience a Reformation crisis. As politique doctrine prescribed, France would allow religious toleration within the realm, while pursuing a secular, sovereign states system, friendly to its security in Europe.

The Reformation had roiled France in the last three decades of the sixteenth century and corresponds closely to France's inclination toward a European system of sovereign states. As in Germany, the Reformation came from below, but it led not to an independent Protestant state but to a sundering civil war, one that in turn evoked the novel, secularizing compromise solution proposed by the politique party. Figure 1 shows this novel pathway of social power.

It was during the early seventeenth century that France first adopted its interest in Westphalia. Although the Habsburg threat dated back to 1519, the French opposition took different forms during different periods. Between 1522 and 1559 France fought to prevent a Habsburg invasion and to expand its buffer zone, but it did not seek to eliminate the empire or the church's temporal powers. Between 1560 and 1598 France imploded in civil war, allowing it scarcely any foreign policy except for a brief defensive war against Spain during the 1590s. In the quarter century after 1598 civil war ceased within the realm, while France remained at peace with Spain and hesitant to intervene in the religious fracases stirring in the empire between 1610 and 1624. It was not until the age of Cardinal Richelieu, a politique who became chief royal adviser in 1624, that France adopted an offensive policy of weakening and then destroying the empire and the Habsburgs and fighting for the sovereign rights of German princes—an interest in European sovereign statehood.[77]

Calvinism evoked this new interest in two ways. Most importantly, the social power of its armed uprising gave rise to the politique ideology, which supported a system of sovereign states. In the 1550s, acting out the first role of ideas, an international network of Calvinist mis-

[77] John Lynch, *Spain under the Habsburgs* (New York: Oxford University Press, 1965); H. G. Koenigsberger and George L. Mosse, *Europe in the Sixteenth Century* (New York: Holt, Rinehart, and Winston, 1968); J. H. Elliott, *Europe Divided, 1559–1598* (London: Collins, 1968), 11–29; Osiander (fn. 13), 27–29; Friedrich Meinecke, *Machiavellianism: The Doctrine of Raison d'Etat and Its Place in Modern History*, trans. Douglas Scott (Boulder, Colo.: Westview Press, 1984), 1; Church (fn. 16), 283–340, 480–82; Richelieu (fn. 16).

sionaries converted French hearers through the sermon, the pamphlet, and favorable circumstances similar to Germany's. As Figure 1 shows, this pathway of conversion was similar to that in Germany, except for the addition of missionaries. Although Protestantism in France was not indigenous, it was successful. Thus, at the onset of the civil wars in the 1560s, 10 percent of the French people and 40 percent of the nobility were Calvinist.[78] The nobles who converted allied immediately with the congregations, some out of conviction, some out of desire for church property, some out of both. Like the Protestant German princes, they confiscated church lands, gained control of local and provincial governments, and then fought off Catholic princes and armies supplied by the congregations.[79]

Between 1562 and 1598 Calvinism spread, Huguenots revolted, a Catholic League formed in response, and eight civil wars ensued. The political costs and incentives facing nobles and kings were radically altered. Nobles could ally with Huguenots or fight them off. If they allied, they could gain wealth, office, and armies, but they would then have to wage war against Catholic nobles and their armies. Kings, almost all of them Catholic, now had to choose between fighting Protestants or tolerating them. Fighting would mean raising armies and using them internally, not abroad; toleration would mean losing the crucial political support of Catholic nobles. After the St. Bartholomew's massacre of the Huguenot leadership of 1572, kings typically issued edicts of toleration only when their resources had been stretched too thin from battling Protestants. Vulnerable paucity, though, was fairly common. Kings relied heavily upon the nobility, to whom they sold offices for revenues and armies, and enjoyed the support of only that faction whose favor they could curry at any one time. Ideas created conflict; conflict sapped the king's resources.

But the same conflict gave rise to the idea—the politique idea—that proved its solution. After thirty years of war had destroyed civic unity and prevented an active foreign policy, the politique option gained in stature and made possible the accession of Henry of Navarre, a quintessential politique. As King Henry IV, he signed the compromise Edict of Nantes, granting Huguenots liberty of worship in specified regions, but not universal toleration. The politics of toleration was not intrinsic to either side's ideals. Although some Calvinists called for religious toleration, many others, including Calvin himself, sought a uniformly Reformed state. In opposition, orthodox Catholics fought to

[78] Porter (fn. 24), 73.
[79] Dickens (fn. 66), 164–87; Elliott (fn. 77), 116–25.

preserve the France of "un roi, une loi, une foi." The politiques still desired religious uniformity yet countenanced toleration in the interests of social peace, to avoid ruin. The foreign policy counterpart of politique doctrine was *raison d'état*, also an expression of toleration and a privileging of state security. In place of lingering fealty to Christendom and the Habsburgs, France would pursue a sovereign states system, an arrangement that would curtail transnational authority and allow it to achieve the balance of power that would best preserve its security. This new stress on security as an end and Westphalia as a means arrived through religious war.[80]

The second way in which Calvinists influenced French policy was through their continued struggle for toleration, since they were discontented with the terms of Nantes. The kings' ability to fight the Habsburgs always depended inversely on Huguenot strength: when Huguenots surged, armies were diverted inward; when Huguenots were quelled (by royal force) or contented (with a new edict of toleration), kings could fight the Habsburgs more vigorously. The defeat of the Habsburgs and the advancement of a system of sovereign states, then, required two conditions: a politique in office and the Huguenots in abeyance. Illustrating the first condition, French foreign policy closely followed the ideology of the most powerful faction in the government during the next half century. When traditional papist *dévots* held power between 1610 and 1624, France remained sympathetic to the empire but militarily uninvolved in the growing continental conflict between Catholics and Protestants. In 1624, though, the politique Cardinal Richelieu took office as chief royal adviser and architect of foreign policy and sought to change course. But before he could commit France irrevocably to the struggle against the Habsburgs, he had to defeat a Huguenot revolt in the mid-1620s. By the 1630s he was ready to pursue war with Habsburg Austria, Spain, and the empire, seeking a comprehensive design for a community of sovereign states to replace the Holy Roman Empire—an aim that no French king had ever before pursued. In France, then, the Reformation's influence was more indirect. It created not a confessional Protestant state, but a politique ideology that called for a system of sovereign states.[81]

[80] On the religious wars in France, see James Westfall Thompson, *The Wars of Religion in France, 1559–1576* (New York: Frederick Ungar, 1909). On the politiques, see Skinner (fn 51), 249–54; and William Farr Church, *Constitutional Thought in Sixteenth-Century France* (Cambridge: Harvard University Press, 1941), 194–271. On Bodin, see Jean Bodin, *On Sovereignty: Four Chapters from Six Books of the Commonwealth* (Cambridge: Cambridge University Press, 1992); and Julian Franklin, *Jean Bodin and the Rise of Absolutist Theory* (Cambridge, U.K.: At the University Press, 1973).

[81] Richelieu (fn. 16); Church (fn. 16).

THE REST OF EUROPE

In detailing the cases of France and Germany, I have shown how Protestant ideas, in their two roles, elicited an interest in Westphalia via two distinct historical pathways. While France was alone in its politique pursuit of sovereign statehood, other polities such as the Netherlands corresponded to the German model of Reformation from below. There, the ruling nobility decided to declare independence from Spain in 1581 only reluctantly and amidst pressure from revolting Calvinists.[82] Transylvania's sixteenth-century Reformation was closer to the German model and allied itself with the anti-imperial forces in the Thirty Years' War.[83]

A third pathway is simpler and can be more easily described: Reformation from above. Here, the monarch, without strong incentives from below, exercised the social power of his position to make the Reformation official in the churches (thereby furthering conversion), seize Catholic lands and powers, and support the Reformation elsewhere in Europe. Even in these cases conversion still occurred across societal sectors, and converts wielded social power from their defiant practice and participation in the armed forces. Figure 1 shows this reciprocal support and empowerment. So monarchs did not act alone, but in comparison with other cases, they exercised much more power and initiative.

Sweden is archetypical. The Lutheran Reformation was officially established by King Gustav Erikkson in the 1520s with the cooperation of, but hardly pressure from, theologians, clergy, and the urban populace. Sweden came to support the sovereignty of German princes shortly thereafter, allying with them diplomatically during the 1540s. A century later, in 1630, Sweden entered the Thirty Years' War against the empire, led by the devout Lutheran king Gustav Adolph.[84] In England, Henry VIII largely initiated the Reformation in the 1530s, acting far more instrumentally than Gustav Adolph would later on, but also benefiting from the societal spread of Protestantism. Though England's own civil wars kept it out of the Thirty Years' War, its foreign policy closely followed confessional lines. As long as its monarchs were Protestant, it generally gave economic aid, arms, and assistance in dy-

[82] Geoffrey Parker, *The Dutch Revolt* (Ithaca, N.Y.: Cornell University Press, 1977), 126–68; idem, *Spain and the Netherlands, 1559–1659* (Glasgow: Fontana Press, 1990), 52–53; Lynch (fn. 77); Geyl (fn. 10); Israel (fn. 10), 106–230.

[83] Parker (fn. 13), 155.

[84] Michael Roberts, *The Early Vasas: A History of Sweden, 1523–1611* (Cambridge: Cambridge University Press, 1968), 68–70; N. K. Anderson, *The Reformation in Scandinavia and the Baltic*, in G. R. Elton, ed., The *New Cambridge Modern History* (Cambridge: Cambridge University Press, 1958), 146–53; Roberts (fn. 16, 1967), 78.

nastic intrigue to its Protestant brethren in the Netherlands and Germany, while it consistently opposed the empire, even defeating the Spanish armada in 1588.[85] Denmark followed a path similar to Sweden's. There, a Lutheran Reformation succeeded by the mid-1530s and was followed quickly by alliances with warring German Protestant princes.[86]

Obversely, in those polities that fought for, or at least aligned with, the empire, the Reformation never managed to make so much as a chink in Catholic piety, practice, and authority. I have noted that in Germany, officially Catholic polities aligned with the empire from the inception of the Reformation up through Westphalia. Most striking is Spain, whose dynastic links with the empire and zealous advocacy of Catholic uniformity in Europe made it the most pro-Catholic polity on the continent. It was the European polity least penetrated by the Reformation. The Reformation mostly failed, too, in the Italian city-states, which remained aligned with the empire during most of the wars of the Reformation. The Italian link is not diminished by the fact that Italian city-states themselves constituted an effective system of sovereign states during the fifteenth century. My argument is not for the necessity of the Reformation in eliciting any system of sovereign states but for its necessity in eliciting Westphalia, a system that the Catholic, post-Renaissance Italian states did virtually nothing to bring about. Finally, Poland and Habsburg Hungary, too, were enduringly Catholic and enduringly loyal to the empire. Poland intervened on the imperial side several times during the Thirty Years' War. Although the Reformation spread significantly within these realms, it never captured the crown or sparked a social crisis like that in France. They complete the continental correlation between Reformation crises and interests in Westphalia.

The Reformation and the Thirty Years' War

The social power of Protestants, then, was coincident with and plausibly connected to the development of an interest in sovereign statehood in all of the polities that together defeated the empire during the Thirty Years' War. But the explanation is incomplete without an account of how these polities actually achieved Westphalia once they had developed an interest in it—through ultimate military victory. Virtually every historian of the Thirty Years' War agrees that ongoing disputes between Catholics and Protestants in Germany and Bohemia were the

[85] Cameron (fn. 51), 381–88.
[86] Ibid., 272–74.

original cause of the war, even as religion became intertwined with other stakes: the ambitions of the military entrepreneur Wallenstein, who amassed immense wealth and lands fighting for the Habsburgs, the fate of Spain as a great power, the sea power of the Netherlands, and the strategic position of Sweden. Most historians would agree, too, that by the last decade of the war religious passions were spent and rulers continued to fight on for wealth and territory. But how states entered into the conflict in the first place is inexplicable apart from the power of Protestant ideas.[87]

Skepticism of Protestantism's Social Power

Structural material explanations challenge the assertion "no Reformation, no Westphalia." The prominence of these explanations makes skepticism of the Reformation significant; that is, it is plausible that ideas were superfluous. But, as I argue, the ultimate inadequacy of these explanations leaves our proposed counterfactual claim about the Reformation intact.[88] The discussion of the German case considered one version of materialism based on skepticism about motives. A more compelling alternative to ideas would not focus on motives, however, but would propose a different sort of social power that better explains the timing and geography of interests in sovereign statehood. It would look instead toward how over the course of centuries states accrued military, economic, and extractive bureaucratic power and saw the growth of their treasuries and their systems of commerce and law. Here, the story of Westphalia is the story of the state, how it grew, and how it triumphed over rival institutions—the Holy Roman Empire, the Hanseatic League, and the Italian city-states.[89] The Peace of Westphalia merely ratified this victory.

Most of the many social scientists who have told the story of the state identify its origins in changes in economics, organization, military technology, or the international distribution of power and in the very activity of war. Here, I do not want to assess the relative merits of these accounts; rather I treat them as a common source of skepticism about the social power of ideas. If the growth of the state turns out wholly to explain the system of sovereign states, then the exact causes of this growth can be debated. But a general statist account can be taken to hold that in a grand historical trajectory running from the High Middle Ages to the eighteenth century, France, England, Sweden, Prussia, other

[87] For histories of the Thirty Years' War, see fn. 13.
[88] See Tetlock and Belkin (fn. 25), 21–25.
[89] Spruyt (fn. 1), 22–33.

German states, and the Netherlands grew into statehood and adopted an interest in a sovereign states system through some combination of material forces. Certainly, there is significant truth in this epic. State growth began well prior to the Reformation, and theorists of state building explain much of it with scant reference to the Reformation. But did the rise of the state power and institutions alone bring about Westphalia? To make this case, such accounts must show that state growth steamed ahead inexorably until the day when polities adopted their interests in a Westphalian system. The Reformation, meanwhile, occurs on the sidelines, its social power redundant or irrelevant.

Returning to Table 1, column 4 shows that the growth of state institutions generally fails to correspond with interests in Westphalia. Over the course of the fifteenth century, admittedly, state institutions across Europe secured stable power over their citizens and notable autonomy from the church, following civil war and conciliar controversy. But the momentum abated. The period between 1500 and the end of the Thirty Years' War is one that several leading early modern historians characterize as one of crisis, war, and monarchical trepidation.[90]

Despite their claims to godlike status, monarchs were continually assaulted by aristocrats who revolted and often resisted the extraction of their revenue. In the first half of the seventeenth century the central state apparatus did expand somewhat in France and Sweden, but revolts and calamity also continued, especially in France. It was not until after the Thirty Years' War that "bureaucratic-military absolutism," epitomized by Louis XIV, finally developed. It was then that the military revolution—itself brought on by the Thirty Years' War—elicited the sharp expansion of the state in France and Prussia and that monarchs came to enjoy the stability and state apparatus which their forebears desired.[91]

But contrast this timing with the development of their interests in a system of sovereign states. In Germany state institutions grew gradually over the sixteenth century, but it was not until the latter half of the seventeenth century, after the Thirty Years' War, that absolute monarchy developed, most prominently in the Brandenburg-Prussia of Frederick the Great. German rulers adopted their interest in Westphalia in

[90] David Kaiser, *Politics and War: European Conflict from Philip II to Hitler* (Cambridge: Harvard University Press, 1990); Geoffrey Parker, *Europe in Crisis, 1598–1648* (Ithaca, N.Y.: Cornell University Press, 1979); Trevor Aston, ed., *Crisis in Europe, 1560–1660: Essays from Past and Present* (London: Routledge and Kegan Paul, 1965); J. H. Elliott, "The Decline of Spain," in Aston; John H. Kautsky, *The Politics of Aristocratic Empires* (Chapel Hill: University of North Carolina Press, 1982), 341–46.

[91] Downing (fn. 47).

the 1530s.[92] Sweden mimicked Germany, achieving the zenith of its power between 1670 and 1718, long after its interest in sovereign statehood had been realized.[93] Similarly, the Netherlands' golden age of commerce and rapid military expansion began in the 1590s, after its declaration of independence.[94]

In France the politique vision of *raison d'état* not only inspired France's contribution to Westphalia but also called for building the crown's powers within the realm. After its civil war died down, the French state grew in its financial and tax-raising apparatus and in its military might. Under Richelieu, its armed forces increased from twelve thousand men in 1629 to two hundred thousand during its intervention in the Thirty Years' War. Yet even then, the crown was weakened by the nobles' constant opposition to reform and taxation and by several uprisings during the 1630s and 1640s. It was not until the reign of Louis XIV and his minister Colbert after Westphalia that France experienced its military revolution and succeeded in raising revenues from the nobility to wage hegemonic war. This was decades after it had developed an interest in a sovereign state system.[95]

The notion that Westphalia is a mere corollary of the Leviathan's bursting forth stumbles in the face of history's timing. Bodies politic were interested in Westphalia before their institutions experienced their dramatic adolescent growth. But the problem of timing is also one of suddenness, not merely sequence. In each case where rulers came to advocate Westphalia, little such talk had been heard even a generation earlier. If the gradual growth of states over centuries explains Westphalia, then why did not their interests also grow gradually, alongside

[92] Ibid.; Gagliardo (fn. 10); Barraclough (fn. 10), 376–80; Carl Cipolla, ed., *The Fontana Economic History of Europe: The Sixteenth and Seventeenth Centuries* (Glasgow: William Collins Sons, 1974); Gerald Strauss, *Law, Resistance, and the State: The Opposition to Roman Law in Reformation Germany* (Princeton: Princeton University Press, 1986); Moeller (fn. 60); Carsten (fn. 75), 165–78.

[93] Parker (fn. 47, 1976), 206; Roberts (fn. 16, 1967), 78.

[94] Parker (fn. 47, 1976), 206; Israel (fn. 10), 106–230; Jan DeVries, *The Dutch Rural Economy in the Golden Age, 1500–1700* (New Haven: Yale University Press, 1974); Henri Pirenne, *Early Democracies in the Low Countries: Urban Society and Political Conflict in the Middle Ages and Renaissance* (New York: Harper and Row, 1963); J. W. Smit, "The Netherlands Revolution," in R. Forster and J. P. Greene, eds., *Preconditions of Revolution in Early Modern Europe* (Baltimore: Johns Hopkins University Press, 1970); Geyl (fn. 10).

[95] Parker (fn. 47, 1976), 206; Richard Bonney, *Political Change in France under Richelieu and Mazarin, 1624–1661* (Oxford: Oxford University Press, 1978); idem, *The King's Debts: Finance and Politics in France, 1589–1661* (Oxford: Oxford University Press, 1981); idem, *Society and Government in France under Richelieu and Mazarin, 1624–1661* (London: Collier-Macmillan, 1988); Church (fn. 16), 81–102, 283–340; Elliott (fn. 77), 11–29; Martin Wolfe, *The Fiscal System of Renaissance France* (New Haven: Yale University Press. 1972); Porter (fn. 24), 74; Victor L. Tapié, *France in the Age of Louis XIII and Richelieu*, trans. D. McN. Lockie (Cambridge: Cambridge University Press, 1984); Robin Briggs, *Early Modern France, 1560–1715* (Oxford: Oxford University Press, 1977); David Buisseret, *Sully and the Growth of Centralized Government in France, 1598–1610* (London: Eyre and Spottiswoode, 1968).

their institutions? This abruptness suggests another form of social power at play.

Even more damning than timing is Spain. Like other European states, Spain gained strength in the fifteenth century, unifying its territory and experiencing, by some measures, the earliest and most rapid growth of any contemporary European state. It expanded its military from 20,000 troops in the 1470s (smaller than England's or France's) to 150,000 in the 1550s (three times that of France), established an overseas empire that fed it hordes of silver and gold, and enlarged its treasury and royal bureaucracy.[96] Yet the Spanish colossus never sought or fought for a Westphalian system of states and was indeed its archopponent, regarding it as heresy. If state growth leads to an interest in Westphalia, then why did not early modern Europe's strongest state develop such an interest? The Spanish state's legitimacy was bound to what remained of medieval Europe: it was pervasively Catholic and the imperial ruler of the Netherlands, its king was the Holy Roman Emperor for the first half of the sixteenth century, and afterward it was still closely tied to the imperially linked Habsburg dynasty. This geographic comparison, too, challenges the story of Westphalia as a mere spinoff of the state's coming of age.[97]

Not only does the state's growth fail to supplant the Reformation as an explanation for Westphalia, but, as this period's embattled politics reveal, this growth itself was deeply shaped by the Reformation. From 1530 to 1648 armed conflict raged continuously at least somewhere in Europe. Most of these conflicts were caused in large part by the dispute between Protestants and Catholics—the most prominent crisis of the period. In most European polities that experienced it, the crisis itself was the chief brake on state growth. In the Netherlands, Prussia, and France it even wrought civil war. It was as a solution to this chaos that leaders embraced the strong secular state. It was also amidst religious war that technological developments led states to expand their militaries, necessitating in turn the need for state institutions to raise the money and the troops to support it. The military revolution, an imminent source of absolutism, arose from wars caused by the Reformation.

Admittedly, factors besides the Reformation helped to bring about Westphalia. During the first half of the seventeenth century Spain suffered bankruptcies and decline, Sweden rose, the Netherlands flour-

[96] Parker (fn. 47, 1976), 206.

[97] J. H. Elliott, *Imperial Spain, 1469–1716* (London: Edward Arnold, 1963); Elliott (fn. 90); Henry Kamen, *Spain 1469–1714: A Society in Conflict* (London: Longman, 1983); J. H. Elliott, *Richelieu and Olivares* (Cambridge: Cambridge University Press, 1984).

ished economically, and the empire became overextended, fighting not only Protestants but also the Turks. But if ideas were not a sole cause, they were more than just another of many, inextricably interwoven causes. The tightness of the geographic and temporal link between ideas and interests in Westphalia, the pathways that expose the cogs and pinions of the influence of ideas, the weak connection between structural material trends and political interests, the partial effect of ideas upon these trends—together, these factors argue that without the Reformation, an international system would have arrived at a very different time, under very different circumstances, in some alternative historical universe.

The Significance of the Argument

The world of sovereign states that emerged at Westphalia was one that realists from Hobbes to Waltz would so enduringly describe, one in which the campaigns of polities to alter religious practice and belief within the borders of other polities came to appear strange, even difficult to imagine. Religion itself, of course, was far from a permanently spent force in international politics; many analysts point to a recent resurgence.[98] Religious conflict appears strange, rather, only in its early modern role, as a major challenge to the legitimacy of the modern international system. Perhaps this strangeness helps to account for the materialist cast of most contemporary social scientific accounts of how this world came to be. But if what I have argued is correct, the realist world itself did not evolve solely from material factors. It was the result of much more: a new conception of humanity's relationship with God.

The case for this new conception has implications for the field of international relations. It adds to a growing corpus of work that asserts the influence of ideas, culture, and norms in shaping international politics. It aims to prove the power of a very large and powerful idea, the Protestant Reformation, and to propose pathways of influence through which the Reformation shaped polities' interests in Westphalia—their adoption of it and their ability to achieve it. The deeper importance of the argument, though, lies in the importance of the event about which it argues. That Westphalia is the founding moment of the international system is old wisdom here renewed. An explanation of this moment is

[98] See Samuel Huntington, *The Clash of Civilizations and the Remaking of World Order* (New York: Simon and Schuster, 1996); Mark Juergensmeyer, *The New Cold War? Religious Nationalism Confronts the Secular State* (Berkeley: University of California Press, 1993); Benjamin Barber, *Jihad vs. McWorld* (New York: Times Books, 1995).

an explanation of the revolution that so enduringly constituted the international social world.

As Westphalia today becomes riven, insights into its origins are all the more momentous. If ideas were a crucial source of the system of sovereign states, they could well be the source of contemporary trends away from sovereignty such as internationally sanctioned intervention and the expansion of the European Union. We can imagine hypotheses: intervention results from the growth of human rights and democracy during the cold war; European integration originated in the post–World War II popularity of European federalism, itself rooted in Christian Democracy and Catholic social thought. Of course, there are material explanations for both trends, so the role of ideas would have to be established. If proven, however, the influence of ideas in both cases would be significant not only for causal debates but also as a case of historical return. In both cases sovereignty becomes restricted on behalf of values that are universal and transcendent— precisely what was lost at the Peace of Westphalia.

Review Article

THE MANY VOICES OF POLITICAL CULTURE
Assessing Different Approaches

By RICHARD W. WILSON

Richard J. Ellis and Michael Thompson, eds. *Culture Matters: Essays in Honor of Aaron Wildavsky*. Boulder, Colo.: Westview Press, 1997, 252 pp.
Michael Gross. *Ethics and Activism: The Theory and Practice of Political Morality*. Cambridge and New York: Cambridge University Press, 1997, 305 pp.
Samuel P. Huntington. *The Clash of Civilizations and the Remaking of World Order*. New York: Simon and Schuster, 1996, 367 pp.
Ronald Inglehart. *Modernization and Postmodernization: Cultural, Economic and Political Change in Forty-three Societies*. Princeton: Princeton University Press, 1997, 453 pp.
David I. Kertzer. *Politics and Symbols: The Italian Communist Party and the Fall of Communism*. New Haven and London: Yale University Press, 1996, 211 pp.

THE popularity of political culture has waxed and waned, yet it remains an enduring feature of political studies. In recent years the appearance of many excellent books and articles has reminded us of the timeless appeal of the subject and of the need in political analysis to account for values and beliefs. To what extent, though, does the current batch of studies in political culture suffer from the difficulties that plagued those of an earlier time? The recent resurgence of interest in political culture suggests the importance of assessing the relative merits of the different approaches that theorists employ.

ESTABLISHING EVALUATIVE CRITERIA

The earliest definitions of political culture noted the embedding of political systems in sets of meanings and purposes, specifically in symbols, myths, beliefs, and values.[1] Pye later enlarged upon this theme, stating

[1] Sidney Verba, "Comparative Political Culture," in Lucian W. Pye and Sidney Verba, eds., *Political Culture and Political Development* (Princeton: Princeton University Press, 1965), 513.

World Politics 52 (January 2000), 246–73

that political culture "encompasses both the political ideals and the operating norms of a polity. . . . [It] is thus the manifestation in aggregate form of the psychological and subjective dimensions of politics." And as he concluded: "A political culture is the product of both the collective history of a political system and the life histories of the members of that system."[2] Political culture thus never addresses individuals or societies singly. It holds that individuals are the fundamental units of societies and does so in a way that embraces the attributes of both personality and social systems. Assumptions emphasize different factors, such as how individuals and/or groups are socialized, how different individuals organize their thinking about rules and norms, how discourse affects the legitimacy of political institutions, how and why individuals orient their thinking and communication in terms of salient myths, rituals, and symbols, and how moral criteria are apprehended and with what consequences for political behavior. Because these assumptions vary in importance depending on intellectual taste and training, it is impossible to speak of a single political culture approach. Rather, different approaches have been utilized to solve different problems and to suit different methodologies.

In seeking to assess the claims of these various approaches one needs a measure against which they can be evaluated. To this end, I suggest beginning with a critical component of all political culture studies—preference formation and, more specifically, how culture constrains preferences and how preferences affect culture. The nature of preferences is an irreducible, necessary, and shared focus of all political culture analyses; the critical factor underlying preference stability is the social and psychological constraints that act upon it. Concomitantly, what happens to preference stability when these constraints become inoperative is also of crucial importance.

One way to conceptualize preference stability is through a modification of the law of relative effect.[3] If we posit that (R) is the rate of expressing a preference, (k) is the maximum possible rate in these terms, (s) is the satisfaction from making choices in terms of this preference orientation, and (sa) is the satisfaction from making choices in terms of alternative preference orientations, then the hypothesis states that

$$R = \frac{k(s)}{s + sa}$$

[2] Lucian W. Pye, "Political Culture," in *International Encyclopedia of the Social Sciences*, vol. 12, ed. David L. Sills (New York: Macmillan and Free Press, 1968), 218.

[3] Peter du Preez, *A Science of Mind: The Quest for Psychological Reality* (London: Academic Press, 1991), 101–2.

This equation tells us that as (s) increases so will (R), dramatically. But to the degree that (sa) is important with respect to (s), then (R) will fail to materialize to any significant degree. As hypothesized here, the probability of (R) being high—that is, the constraints that maintain a significant degree of (s) are high—is related to the following factors:

— the degree of congruence in the cognitive and affective valences of society's members
— the degree to which the norms and symbolic referents that coordinate social interaction are based on mutually reinforcing definitions and are widely accepted as legitimate
— the degree of congruence in individual and social interpretations of reality
— and, at the individual level, the degree of consistency in response repertoires

Such constraints obviously have an impact on the value of (k), the maximum possible rate of expressing a particular preference. When these constraints are weak or nonexistent with respect to (s) and/or when entirely different criteria become significant as constraints on preferences, the probability of (sa) being large relative to (s) is likely. In this circumstance social (and cultural) disorganization will be marked, as groups and individuals struggle to maximize alternative preference orientations through peaceful discourse, street demonstrations, revolution, and so on.

This essay begins by assessing five different books, each representative of a political culture approach—the five approaches that largely define the theoretical space of current political culture studies. In exploring how (s) and (sa) are conceptualized in each approach, the relationship between them, and the affect that this has upon (R), it becomes clear that while each approach has its strengths, not all are equally effective in explaining rates of preference formation (R). This is a key point, for if political culture is to have utility in political analysis it must be able to explain why and when people choose particular courses of action over others and the likelihood that these choices will be widely shared. The second part of the essay considers the difficulty political culture theorists have in confronting the relationship between preference formation and political cultural norms. Finally, the essay briefly highlights some promising avenues for future research.

Assessing Political Culture Approaches

To compare political culture approaches, one needs to assess how they identify and measure the relative magnitudes of (s) (the satisfaction from making choices in terms of a particular preference orientation)

and (sa) (the satisfaction from making choices in terms of an alternative preference orientation). This yields a critical measure of theoretical adequacy: how well an approach can determine (s) and (sa), identify constraints on their formation, and establish a plausible connection between constraints and satisfaction.

Each of the books reviewed deals broadly with the question of political transformation, utilizing different criteria to explore this phenomenon. In evaluating the approaches attention is paid to how alternative ways of assessing preferences affect our understanding of political change.

Political Culture Approaches in the Anthropological Tradition

THE HERMENEUTIC OR INTERPRETATIVE APPROACH

One powerful framework of explanation is grounded in the anthropological tradition of fieldwork and immersion in community life.[4] Many political scientists consider these studies to be especially appropriate for areas lacking survey data and for studies of local communities; this type of work has produced many useful concepts and midrange theories that are as applicable to national polities as they are to local levels. The focus on the symbolic dimension of politics, including the role of myth and ritual, has been applied to analyze nineteenth-century Bali and to explain changes in societies as diverse as contemporary Poland and Israel.[5] Derived from both a Weberian concern with the meaning of social action and a phenomenological focus on the socially constructed nature of knowledge and society, culture in this approach is defined as the collective meanings that groups create, share, and symbolically express. There is commonly a myth of origin, which relates how a people came to be or how a nation was created. Ideologies, the more particular interpretations of these wider cultural themes, tend to be more concious and rationalized versions that give legitimacy to the claims of

[4] See, as examples, Lowell Dittmer, "Political Culture and Political Symbolism: Toward a Theoretical Synthesis," *World Politics* 29 (July 1977); Murray Edelman, *Constructing the Political Spectacle* (Chicago: University of Chicago Press, 1988); Laura Desfor Edles, *Symbol and Ritual in the New Spain: The Transition to Democracy after Franco* (Cambridge: Cambridge University Press, 1998); James W. Fernandez, *Persuasions and Performances* (Bloomington: Indiana University Press, 1986); Bruce Kapferer, *Legends of People, Myths of State* (Washington, D.C.: Smithsonian, 1988); Sally F. Moore and Barbara G. Myerhoff, *Secular Ritual* (Assen, Netherlands: Van Gorcum, 1977); Sherry B. Ortner, *High Religion* (Princeton: Princeton University Press, 1989).

[5] Myron J. Aronoff, *Power and Ritual in the Israel Labor Party* (Armonk, N.Y.: M. E. Sharpe, 1992); Clifford Geertz, *Negara: The Theatre State in Nineteenth-Century Bali* (Princeton: Princeton University Press, 1980); and Jan Kubik, *The Power of Symbols against the Symbols of Power: The Rise of Solidarity and the Fall of State-Socialism in Poland* (University Park: Pennsylvania State University Press, 1993).

different groups.[6] Although studies using this approach employ a variety of methodological techniques, participant observation remains a central one, and surveys are generally supplementary. As a consequence, the approach interrelates well with studies that seek explanations for community fragmentation or cohesion in situations where dramatic macrolevel political change is occurring.

David I. Kertzer in *Politics and Symbols* develops an interesting and informative account of the demise of the Italian Communist Party. So swiftly did this occur that many of the party faithful were shocked and dismayed. Although as early as the 1960s trust in the USSR as the leader of the world's proletariat had eroded, by the mid-1970s both Italians and foreigners were predicting that the party might even come to hold a majority in the national government. Nonetheless, the party's momentum began to slacken, such that by 1989 events in Eastern and Central Europe accelerated the belief of the party secretary, Achille Occhetto, and of others that bold action to transform the party was required. This transpired in November 1989, with Occhetto's announcement of the *svolta*, or turnabout—the need to change the name of the party and create a new-style leftist political entity.

Kertzer spent a year in the working-class quarter of Bologna talking with those involved, attending meetings, visiting cafes, and participating in protest meetings and rallies. Occhetto's statement had clearly provoked a crisis resulting in huge defections from the party and a drop in the share of the party's vote, from 30 percent in 1985 to 24 percent in 1990 (p. 14). The name of the party and the use of its emblems and symbols were hotly contested by a powerful secessionist group, and there soon developed a battle over the party's history, its myths, and the construction of a new postcommunist identity. This conflict was the basis for Kertzer's "concern . . . with the role of symbolism and the manipulation of history in the Communist struggle" (p. 15), and he turned his focus to "the evolution of the party's Communist identity" (p. 15).

The twin processes of mythologization and ritualization, says Kertzer, create history. The party's identity was bound up in a mythologized past associated closely with the resistance against fascism. The events of this struggle, not always depicted by the faithful with accuracy, were used by communists over the years to construct the themes of an evil conspiracy, the existence of a savior, and the coming of a golden age. When the members of the party discovered that they no longer had a name (that is, Communist), the situation signified as well the loss

[6] Aronoff, *Israeli Visions and Divisions: Cultural Change and Political Conflict* (New Brunswick, N.J.: Transaction Publishers, 1989), xv–xvi.

of the symbols, myths, and rituals associated with that name, for all were inextricably intertwined. Notwithstanding, in the ensuing political struggle both Occhetto and his opponents drew on the symbolism of the party while redefining both the party's symbols and its history to suit tactical needs.

The analysis of symbols is central for Kertzer because he believes that politics is at root a symbolic activity. This led him to explore how important symbols (for example, the red flag and revolutionary martyrs) are constructed and altered and how some symbols become legitimate and others are rendered illegitimate. Crucial to understanding the importance of symbols is the nature of ritual defined as "any symbolic behavior that is socially standardized, repetitive, and meaningful. Indeed, it is through ritual that symbols come to be defined, diffused, and energized" (p. 123). Prior to the *svolta* it was ritual that had promoted solidarity by legitimating "a symbolic descent from common ancestors whose mythic status [was] promulgated through rites" (p. 125). As a consequence, standing at the heart of the conflict over the transformation of the party was a struggle over the symbols on which rites were based.

Kertzer claims that "the emotional dimension is central" (p. 156), but nowhere is there a discussion of the psychological mechanisms that link emotions to preferences in a way that can help explain why preferences do or do not change. The result is an interesting account, but aside from the author's assertions no convincing argument is advanced about how "emotions" stimulate the process of identification with myths, symbols, and rituals. Missing is the critical why. Instead, one comes away with a plausible but nevertheless fundamentally narrative explanation of the unique features surrounding the transformation of the Italian Communist Party.

The hermeneutic approach looks to uncover constraint, in the form of myth, ritual, and discourse, largely through immersion in community life (where possible) and by "thick" description, relying heavily on semiotic analysis. There is no explicit psychological dimension. Rather, causal relationships are validated through observation of repetitive patterns: habits and habit complexes at the individual level and customs and custom complexes at the social level. Conceptualizing (s), the satisfaction derived from making choices in terms of a particular preference orientation, and (sa), the satisfaction gained from making choices in terms of an alternative orientation, is done largely by reference to variables at the social level (for example, shared adherence to a common myth of origin) and is achieved by expert interpretation based on ob-

servations of individual and mass behavior in actual social contexts. Satisfaction is inferred from public activity while the degree of constraint, linked to (s) and (sa) respectively, is determined by the inferred strength of enduring predispositions, loyalties, and so on regarding political life.

THE "CULTURE THEORY" APPROACH

The culture theory approach has evolved from the anthropological tradition, in particular, from the work of Mary Douglas.[7] It explicitly eschews direct reference to individual psychological variables and focuses instead on typological distinctions between subgroups said to exist in every polity. These groups are differentiated by two criteria: (1) the extent to which individual behavior within the group is socially prescribed (termed "grid") and (2) the degree to which persons are locked into membership in their group (termed "group"). Based on the relative strength of these criteria, ideal subcultures are identified according to behavior patterns that vary from *apathetic* (fatalism: high grid, low group), to intense concern for *hierarchy* (collectivism: high grid, high group), to *competitive* (individualism: low grid, low group), to *equalitarian* (egalitarianism: high group, low grid), and, finally, to *autonomous* (retreatism: neutral group and grid). The importance of actual subcultures in any society is ascertained by research into the prevalence of these behavior patterns, with the ways that subcultures interact providing the theoretical basis for understanding the political life of a society. In studies of political change this approach is especially well suited for analyses of contestation. Coalitions and more loosely knit groupings are identified in ways that differ from standard sociological (for example, class) or psychological (for example, relative deprivation) formulations.

Through its impossibility theorem culture theory posits that in any society there are exactly these five theoretically identified social relationship patterns—no more and no less. As individuals transact with each other, they come to know more clearly what they want and choose the pattern that is most congenial for them; subsequent preferences are then derived from and constrained by the particular social relationship

[7] See, as examples, Mary Douglas, ed., *Essays in the Sociology of Perception* (Boston: Routledge and Kegan Paul, 1982); Mary Douglas and Aaron Wildavsky, *Risk and Culture* (Berkeley: University of California Press, 1982); Richard J. Ellis, *American Political Cultures* (New York and Oxford: Oxford University Press, 1993); Michael Thompson, Richard Ellis, and Aaron Wildavsky, *Cultural Theory* (Boulder, Colo.: Westview, 1990); Aaron Wildavsky, "Choosing Preferences by Constructing Institutions: A Cultural Theory of Preferences," *American Political Science Review* 81, no. 1 (1987); idem, "Indispensable Framework or Just Another Ideology? Prisoner's Dilemma as an Antihierarchical Game," *Rationality and Society* 4, no. 1 (1992).

pattern a person has identified with. Even as each way of life is dynamically stabilized, however, individuals may shift away from allegiance to one pattern as a consequence of a cumulative mismatch between the promise and the performance of a particular relationship pattern. Microculture is thus the type of choice (that is, hierarchical, equalitarian, and so on) that is associated with a particular relationship pattern, whereas macroculture is the relative strength of the patterns.

Toward the end of his life Aaron Wildavsky became a leading proponent of culture theory, arguing against the assumption that history is merely narrative and that culture is simply national, religious, ethnic, racial, or organizational distinctions that defy common measures. *Culture Matters*, edited by Richard J. Ellis and Michael Thompson, is a collection of essays written in honor of Wildavsky, although only half of them are related to culture theory.

Wildavsky believed that a theory should be judged not against an impossible ideal but in comparison with rival theories. The article by Gunnar Grendstad and Per Selle, "Cultural Theory, Postmaterialism, and Environmental Attitudes," does just that, attempting to compare culture theory with alternative approaches. Employing an elaborate test of a public opinion survey conducted in Norway, they found that Ronald Inglehart's[8] hypothesized connection between environmentalism and postmaterialism was a move in the right direction but that the strength of the relationship was weak to nonexistent. In contradistinction they note that the egalitarian subculture, when operationalized, is able to explain far more variation in environmental attitudes (pp. 162–63). Indeed, materialism-postmaterialism were the weakest predictors of environmental attitudes, leading Grendstad and Selle to conclude that Inglehart was either wrong about the relationships or that his measures of materialism and postmaterialism are inadequate (p. 165). While this is but another of the numerous challenges posed to Inglehart and rebutted by him, it does provide support for the predictive capability of culture theory.

Of critical importance in assessing culture theory's claims is the adequacy of its explanation of political change. Charles Lockhart deals with this issue in "Political Culture and Political Change," maintaining that cultural biases are vulnerable to dissolution. According to the theory's postulates, all people must be associated with a subculture of their choosing. As circumstances change, however, a cumulative mismatch

[8] Ronald Inglehart, "Post-Materialism in an Environment of Insecurity," *American Political Science Review* 75, no. 4 (1981).

between performance and expectations may cause the adherents of any particular cultural bias to alter their preferences. A theory of "surprise" (akin to cognitive dissonance) underlies the shift from one bias to another as changing circumstances foster new institutional relationships, which in turn reshape perceptions. Nonetheless, despite the fact that change is said to be inherent and that shifts may alter the relative influence of rival subcultures (important, for example, for understanding the rapid transformation in the balance of cultural biases in Germany following World War II), fundamental changes in cultural commitments are infrequent. The adherents of a subculture are more likely to change institutional arrangements to support their biases than to do the converse (p. 95).

Change is thus conceptualized by culture theorists as "inherent in the different competencies and biases of different cultures" (p. 99). Nowhere, however, is there a good psychological explanation for why people select a particular bias in the first instance or what the intensity of their attachment to that bias is.[9] Rather, preferences are said to derive from choosing a particular way of relating with others. The analyst determines the distribution in society among preference orientations by identifying patterns of social interaction. The strength of the approach therefore is in its division of individuals into subcultural types, which are then used to explain preferences, while its weakness is the lack of a psychological mechanism to explain why particular preference orientations were chosen to begin with. A more complete explanation can be found only by drawing upon other general theories that divide people into cultural types.[10]

Despite the feeble psychological dimension of culture theory, experts are able, through immersion in community life, to identify at the macrolevel the preference orientations associated with (s) and (sa), assess their relative strength, and determine qualitatively the relative weighting of a dominant (s) and a "counterhegemonic" (sa). Changes in the relative strength of (s) and (sa) are revealed by shifts in their salience. While the theory claims more than it delivers, it minimizes the ad hocism and problems of verification that are the weaknesses of the hermeneutic approach. It directs research toward the discovery and relative weighting of preferences that are theoretically postulated. By so

[9] For an attempt to incorporate psychological variables, see Karl Dake, "Orienting Dispositions in the Perception of Risk: An Analysis of Contemporary Worldviews and Cultural Biases," *Journal of Cross-Cultural Psychology* 22, no. 1 (1991).

[10] Alan Page Fiske, *Structures of Social Life: The Four Elementary Forms of Human Relations* (New York: Free Press, 1991).

doing it clarifies the dilemmas that are embedded within any overall cultural configuration and theoretically accounts for change. Understanding the macropolitical changes that occur as a consequence of preference alterations at the microlevel is as yet poorly conceptualized, however.

POLITICAL CULTURE IN THE PSYCHOLOGICAL TRADITION

THE SOCIAL CHARACTER APPROACH

An alternative interpretive approach traces its roots to national character studies and to anthropological studies in culture and personality.[11] Personality variables, drawn frequently but not invariably from Freudian formulations, are utilized as the major means for typing significant value orientations. Thus, societies are described in terms of the unique ways that their members view authority relations, are committed to particular religious/ideological views, fear social disorder, are excessively dependent, and so forth. Case-study analysis of social norms and behavior reveals contextually how the tensions that are associated with particular orientations are conceptualized and resolved. The approach thus supplements contemporary efforts to unravel why regimes make the transition to (or from) democracy and why populations suddenly explode into revolutionary activity.

It is false, immoral, and dangerous to believe in the universality of Western culture, argues Samuel P. Huntington in *The Clash of Civilizations and the Remaking of World Order* (p. 310). This central thesis meshes with his primary theme "that culture and cultural identities, which at the broadest level are civilization identities, are shaping the patterns of cohesion, disintegration, and conflict in the post–Cold War world" (p. 20). According to Huntington, there are currently seven or eight major civilizations. By studying the interactions among them, he believes, researchers can produce a simple map for understanding the international politics of the late twentieth century.

For Huntington the "central elements of any culture or civilization are language and religion" (p. 59), with language second only to religion

[11] See, as examples, Edward C. Banfield, *The Moral Basis of a Backward Society* (New York: Free ress, 1967); Daniel Bell, *The Cultural Contradictions of Capitalism* (New York: Basic Books, 1976); aniel Goldhagen, *Hitler's Willing Executioners: Ordinary Germans and the Holocaust* (New York: Alfred . Knopf, 1996); Alex Inkeles, *Exploring Individual Modernity* (New York: Columbia University Press, 983); Thomas A. Metzger, *Escape from Predicament: Neo-Confucianism and China's Evolving Political ulture* (New York: Columbia University Press, 1977); Lucian W. Pye, *The Mandarin and the Cadre: hina's Political Cultures* (Ann Arbor: Center for Chinese Studies, University of Michigan, 1988); loyd I. Rudolph and Susanne Hoeber Rudolph, *The Modernity of Tradition* (Chicago: University of hicago Press, 1967); and Max Weber, *The Protestant Ethic and the Spirit of Capitalism*, trans. Talcott arsons (Los Angeles: Roxbury Publishing, 1996).

as a defining feature (p. 70). The relationship of religion to culture is asserted repeatedly (pp. 42, 47, 254), although the connection is never proved and subtle nuances are entirely dispensed with. Thus, Western Christianity confronts Greek Orthodoxy, Islam, Hinduism, Confucianism, and Buddhism, as well as curious hybrids—the civilizations of Japan, Latin America, and Africa. So complete at times is the identity drawn between religion and culture that the very term culture often seems superfluous.

Huntington has an exceptionally gloomy view of the world. Within the United States he sees fundamental conflict between the multiculturalists and the defenders of Western civilization (p. 307). Outside the United States, according to Huntington, the power of non-Western civilizations is increasing while that of the West is declining. He contends that the ability of the West to impose its ideas about human rights and democracy is fading, although the evidence on this point is certainly controversial. The most troubling aspect of Huntington's worldview is the psychology that he imputes to the members of what are presumably irreconcilable civilizations. Conflict, he says, is ubiquitous, for "it is human to hate. For self-definition and motivation people need enemies: competitors in business, rivals in achievement, opponents in politics. They naturally distrust and see as threats those who are different and have the capability to harm them" (p. 130). Thus he concludes that if Chinese power is expanding and that of India is growing then "conflict seems highly probable" (p. 244).

Nothing appears as menacing for the West, however, as the looming confrontation with the Islamic world. In the 1990s nine of twelve intercivilizational wars (as of the time of writing) were between Muslims and non-Muslims, a fact that presumably bodes ill for any peaceful accommodation among civilizations. This, unfortunately, is not the worst of it, in Huntington's view: there is now a Confucian-Islamic connection that threatens ultimately to sweep the West to its doom.

Much of this account borders on one-sided conjecture. No one would dispute the often thorny relations between the United States and China or the difficulties that Muslims and non-Muslims alike have faced with Islamic fundamentalists. Yet as Huntington's critics have pointed out, he ignores the huge differences among peoples within his civilizations and the fact that to greater or lesser degrees all societies are experiencing similar developmental pressures. Huntington, however, believes that there is little or no evidence to support the idea of significant convergence in values and beliefs among the world's populations (p. 59)—this despite the fact that a walk down a major street in any

world city reveals that in the conflict between "Jihad and McWorld" it is by no means clear that Jihad is winning.[12]

Frightened and angry people populate Huntington's civilizations. They are referred to en masse as Chinese, Muslims, and so on, and their preferences are similarly imposed upon them. Huntington knows there are significant internal differences, but his definition of culture as derived basically from religion and language allows little room for nuanced interpretation or for understanding cultural change, as the latter would necessarily be tied to the dynamics of linguistic and religious change. Indeed, given his focus on conflict and his perception of the sinister nature of religious fundamentalism, there is little possibility for any significant cultural change that would ameliorate the clash of civilizations.

This approach links the anthropological and psychological approaches but shifts its emphasis from symbolic interpretation and the analysis of discourse to the psychology of learned behavior. Populations are said to possess modal and enduring personality characteristics (that is, psychological constraints) that are imparted from one generation to the next via political socialization. Individuals enter political life as members of families, clans, religious groups, ethnic blocs, and so on. In some formulations group patterns with deep historical roots are said to uniquely mold values and behavior, as seen, for example, in the persistence of civic virtues in northern Italy.[13] Specifying group membership and loyalties is thus critical for explaining the prevalence of democratic values, religious affiliations, ethnic identifications, and so forth. The approach carefully delineates the reasons for the satisfaction derived from preferences in terms of a prevailing pattern of constraints (s) but is less successful in providing a psychological explanation for the emergence of alternative preferences (sa). As a consequence, (R) is largely inferred. Cultural ambiguity and conflict are invariably explained descriptively from analysis of group behavior. Given the methodological emphasis on insightful interpretation, empirical comparisons of different societies are virtually impossible.

THE SOCIAL LEARNING APPROACH

In this approach values and beliefs are ascertained from survey data that have been subjected to various forms of statistical manipulation.[14]

[12] Benjamin R. Barber, *Jihad vs. McWorld* (New York: Random House, 1995).

[13] Robert D. Putnam, *Making Democracy Work: Civic Traditions in Modern Italy* (Princeton: Princeton University Press, 1993).

[14] See, as examples, Gabriel A. Almond and Sidney Verba, eds., *The Civic Culture Revisited* (Boston: Little, Brown, 1980); Ronald Inglehart, *Culture Shift in Advanced Industrial Society* (Princeton: Prince-

These analyses, grounded in psychological theories of social learni consider socialization in particular and the influence of social life general to be critical determinants of values and behavior. The best these studies have focused on societies where systematic survey researc is possible. To simplify somewhat, the psychological model that in forms these analyses is one in which the values and beliefs of a populac are assumed to reflect the influences of socialization to a significant de- gree. It is thus the environment—social and physical—that conditions individual responses. Given the importance of socialization, it is necessary to conduct longitudinal studies to demonstrate how intergenerational differences in preferences correlate with changes in external influences. Cross-national surveys of values and beliefs are based on the same assumption of differences among societies in their historical and contemporary conditions. By identifying long-term transformations in preference patterns, this approach fits well with studies that seek to make sense of how technological changes, increasing educational levels, and so forth affect social and political life.

Wildavsky puzzled over the rise of egalitarianism in the United States. Inglehart provides a stunning explanation in *Modernization and Postmodernization*. He argues that economic development, cultural change, and political change go together in predictable ways to create fundamental alterations in value systems and basic changes in the hierarchical, bureaucratic organizations that have been essential for the creation of modern society. For Inglehart culture is "the subjective component of a society's equipment for coping with its environment" (p. 55). The interaction between culture and environment may yield profound changes in the way that people view the world. His book is the latest of his works to show how a shift from modernization to postmodernization has been mirrored by an increase in postmaterialist values.

A shift in priorities from materialist to postmaterialist values involves the steady diminution of those values that played a key role in industrialization, such as economic achievement and rationality, and an increasing emphasis on self-expression and quality of life. Change is driven by rising feelings of security. While the shift is not linear, and while the materialist-postmaterialist dimension is only one along which change is measured, the tendency is clear. As Inglehart says: "The linkage between economic growth rates and the percentage of values and

ton University Press, 1989); Herbert McClosky and John Zaller, *The American Ethos* (Cambridge: Harvard University Press, 1984); Harry C. Triandis, "The Psychological Measurement of Cultural Syndromes," *American Psychologist* 51, no. 4 (1996); and Sidney Verba et al., *Elites and the Idea of Equality* (Cambridge: Harvard University Press, 1987).

norms shifting in the Postmodern direction is significant at the .01 level" (p. 321). While a number of different values are involved, it is movement from traditional authority to secular rational authority and from survival values to well-being values that is *most significant. Indeed,* these two dimensions alone account for 51 percent of cross-national variation (p. 82).

Inglehart's findings are based on tests of two hypotheses: (1) a socialization (that is, social learning) hypothesis that states that change will not occur quickly since the basic values that individuals hold reflect the conditions that existed in their preadult years and (2) a scarcity hypothesis that states that because people value what is in short supply, increasing prosperity will bring a concomitant, predictable shift from materialist to postmaterialist values. Because experiencing prosperity is a formative influence, the emergence in a society of postmaterialist values will occur with generational change; its increasing salience can be measured by analyzing different birth cohorts over time. Inglehart has done this on a massive scale beginning nearly three decades ago with samples from a small number of Western European countries and expanding in this work to sixty thousand respondents from forty-three countries representing 70 percent of the world's population and with per capita incomes ranging from $300 to $30,000.

One of Inglehart's most provocative findings mirrors Huntington's identification of civilizations but with vastly different implications. The data reveal that forty of the forty-three societies fall into coherent, historically meaningful clusters on the most significant dimensions such that it is possible to identify a Northern European zone, a Confucian zone, a Latin American zone, and so on. Differences reflect levels of economic development, with the values of richer countries varying systematically from those of poorer countries. Of even greater importance, the ratio of postmaterialists to materialists rises over time in virtually every case. Indeed, advanced industrial societies with widely varying cultural traditions and different economic levels are shown to be experiencing similar changes in political, economic, religious, sexual, and gender norms (p. 49).

Inglehart has offered a new take on modernization theory—one that is not linear, deterministic, or tied to Westernization or democratic theory. He avoids impressionistic historical evidence and relies instead on tightly constructed surveys for data that reveal widespread shifts in preference orientations and societal changes in the configuration of preferences. Not only do younger birth cohorts have a greater preponderance of postmaterialists than older ones, but there is also a notable transformation in in-

stitutional forms. Economic development leads to structural changes (for example, urbanization and mass education) and to attitudinal changes as well (for example, greater distrust and skepticism regarding hierarchical institutions). In virtually all societies where postmaterialism is significantly in evidence, political action that challenges elites increases to the point where a strong positive correlation is found between the postmodernization dimension and democracy (p. 104).

The assumptions of the social learning approach are clear. The use of statistical controls permits replicability and comparison on dimensions independent of expert interpretation. Causation is additive and not conjunctural; estimations can be made of the separate contribution of each independent variable allowing probabilistic statements about a variable's net effect in a wide variety of settings. Alongside these strengths, however, are some significant weaknesses. Surveys are expensive and time-consuming and often difficult to conduct in the very societies where it would be desirable to understand the political culture, namely, societies in crisis (for example, where revolutions are in progress) and countries with closed, authoritarian regimes.

The most difficult problem relates to the conception of culture. In Inglehart's work as the satisfaction from espousing a postmaterialist orientation grows (s), and as the satisfaction from holding a materialist orientation decreases (sa), the rate (R) of expressing postmaterialist preferences increases. As postmaterialism becomes a significant orientation, support for democratic institutions and democratic norms increases. The feedback loop is thus closed; normative prescriptions and preference orientations are reciprocally related. However, the postmodernization cultural dimension needs to be more fully elaborated and greater care needs to be given to explaining the definition of culture as a "subjective component" (p. 55). For instance, interpersonal trust and subjective well-being are noted as the two most prominent cultural values that sustain democratic institutions (pp. 194, 197), although both are more properly individual psychological dispositions than they are normative injunctions. As a consequence, the causal significance of culture in the social learning hypothesis, independent of feedback, is unclear.

THE COGNITIVE EPISTEMOLOGICAL APPROACH

The study of political culture has recently incorporated Piagetian theory in its theoretical design.[15] These analyses differ from social charac-

[15] See, as examples, Dan Candee, "Ego Developmental Aspects of New Left Ideology," *Journal of Personality and Social Psychology* 30, no. 5 (1974); Stephen Chilton, *Defining Political Development* (Boulder, Colo.: Lynne Rienner, 1988); Nicholas Emler, Stanley Renwick, and Bernadette Malone,

:er studies in the way they conceptualize value orientations. They also liffer from studies utilizing survey research in the reduced importance :hey impute to social influences, unmediated by individual psychologi- :al variables, as determinants of values and beliefs. In this perspective ndividual and social life are conceived as developmentally and recipro- :ally organized, if not necessarily synchronous. There is a continuous nteraction between individual- and social-level constructions of mean- ng, between the ways that individuals think and the norms of the po- itical culture. Age-related changes in the organization of thinking ccur at the individual level, and these lead to increasingly abstract and :omplex conceptions about obedience, obligations, relations with au- :hority, reciprocity, and benevolence. And at the social level increasingly :omplex normative prescriptions evolve to minimize the transaction :osts involved with maintaining status hierarchies and to provide a moral underpinning for social solidarity. Because reasoning about so- cial issues inevitably involves the invocation of moral criteria, it is moral reasoning that especially influences the norms of the political culture. The consequence is a specification of preference constraint that is de- pendent on both an individual's stage of moral development and the nature of normative prescriptions.

Michael Gross in *Ethics and Activism* confronts a paradox: what we want in a democratic political culture are individuals with considered moral judgment who act judiciously and who resolve disputes in a peace- ful, institutionalized manner, but what we get may be something quite different. For as Gross points out, activism to foster a democratic politi- cal culture may involve groups whose incentive structures are highly parochial and whose members are motivated by materialistic goals.

Gross begins by asking what best anchors democratic integrity in ethical activism. Is it strong political morality, exemplified in the works of Locke, Mill, Rawls, and Habermas? Or is it weak political morality, as found in the work of Madison. Strong political morality states that collective action to remedy injustice requires that citizens have both moral insight and the cognitive capacity to recognize injustice. In pos- session of these qualities, they will feel an obligation to become active defenders of democratic norms. Gross concludes, however, that moral

---

"The Relationship between Moral Reasoning and Political Orientation," *Journal of Personality and So- cial Psychology* 45, no. 5 (1983); Shawn W. Rosenberg, *Reason, Ideology and Politics* (Princeton; Prince- ton University Press, 1988); Shawn W. Rosenberg, Dana Ward, and Stephen Chilton, *Political Reasoning and Cognition; A Piagetian View* (Durham, N.C.: Duke University Press, 1988); John R. Snarey, "Cross-Cultural Universality of Social-Moral Development: A Critical Review of Kohlbergian Research," *Psychological Bulletin* 97, no. 2 (1985); and Richard W. Wilson, *Compliance Ideologies: Re- thinking Political Culture* (New York: Cambridge University Press, 1992).

behavior and moral-political behavior are two different things. As a consequence, it is weak political morality, with its emphasis on benign regulation and nonviolent participation, that best anchors democratic integrity by stressing limited social altruism and a relatively narrow concern for the common good.

Gross supports this conclusion with three case studies backed by extensive empirical evidence. The first of these involves groups that rescued Jews during the Second World War. He found that among those sheltering Jews moral competence did have an effect on rescue activity but only when joined with strong parochial motivations or weak postconventional moral preferences (p. 149). In his second case, an analysis of pro-life and pro-choice abortion activists, he found that normative incentives had a weak effect on activism while material and solidarity incentives had a strong effect (p. 173). He thus concluded that "without exception, no measure of moral development has any significant effect on any form of activism" (p. 179). In his last case study, of peace activism in Israel, he came to conclusions that were similar to those drawn in the abortion study. What emerges from analysis of the cases, therefore, is a paradox: the morally competent are politically incompetent. Successful collective action rests on parochial norms that are inherent in small and interdependent communities.

Specifically Gross found that "moral outrage does not mesh well with effective political action" (p. 115) in part because moral judgment is not always a necessary condition for moral-political action. The most effective moral motivations, in fact, are anchored in narrow perceptions of altruism, moral duty, and fairness. The most successful collective action is cultivated in small groups that have well-developed social and organizational infrastructures with intense parochial affiliations. In such groups conventional moral reasoners, who in any case are always a majority in societies and whose moral outlooks stress the correctness of group norms, are especially well integrated cognitively. Gross believes the chief obstacle to successful activism is not insufficient moral energy but free-rider problems. As a consequence, organizational incentives that tap rational, self-interested concerns within the context of tight social networks are required to overcome collective action problems. The successful pursuit of moral claims occurs when effective organization is linked to rational and solidarity incentives.

Do these conclusions mean that morality has no place? Not at all. Action proceeds from a perception that moral interests are threatened. The material and solidarity benefits that are crucial for collective action

would be useless without a clearly delineated moral community. Normative incentives, moreover, are the strongest reason for joining a group, although it is personal identity norms (such as guilt and self-interest) that actually sustain activism. Interestingly, because moral incentives are amenable to a utilitarian calculus (solidarity and material benefits), both cognitive psychology and rational choice offer insights into understanding political morality and behavior. Cognitive epistemology sets forth the stages of moral development that range from self-interest to a Kantian perspective; it explains why people cleave to particular moral orientations. Collective action and incentive theory explain how individuals utilize their convictions in politics. "Ultimately," says Gross, "the two psychological theories can be combined into a single, more predictive model to investigate ethics and political activism" (p. 65).

The cognitive epistemological approach provides a coherent explanation of individual-level reasoning and its link to preference formation, but it is theoretically complex and highly abstract, positing that it is the moral development of individuals that changes the relationship between (s) and (sa). At a given point in the maturation process an individual reasons in a certain manner. Only with maturation does there develop a new pattern of reasoning (sa) that reduces the probability of choices in terms of a developmentally lower stage of reasoning (s). The relationship of (s) to (sa) thus shifts, seriatim, as new and more complex structures of reasoning evolve. However, given that people are always at different stages of development, determining how and in what way preferences affect the norms of political cultures is still a matter of conjecture. Michael Gross has provided an intriguing answer to this problem.

Preferences and Political Culture

It is surprising that political culture theorists have had such difficulty confronting the question of preference formation. The answer, I believe, lies in a tendency to conflate political culture and preferences—to assume, on the one hand, that an explanation of political culture reveals the preferences of those who live within that culture or, on the other hand, that an explanation of preferences and preference formation is sufficient to reveal the political culture. We should ask, therefore, what a political culture is and what its relationship is to preference formation. Any understanding of political culture must first ascertain what the function of political culture is within a social system. Such an ex-

planation necessarily takes us beyond the description of a political culture as a distribution of values or as meanings embedded in myths, symbols, and rituals.

In the most general sense political cultures are socially constructed normative systems that are the product of both social (for example, rules that coordinate role relationships within organizations) and psychological (for example, the preferences of individuals) influences but are not reducible to either. They have prescriptive qualities that stipulate not only desired ends but also appropriate ways to achieve those ends. The norms are not coterminous with legal codes, although they often overlap. They are manifested in a variety of ways, including symbolic representations and the content of hegemonic discourse, especially of socialization. Preferences, by contrast, are ends desired by individuals. They vary widely depending on differences in individual experience and psychology. The relationship between the two levels is interactive in the sense that individual preferences continuously affect the normative order while the norms themselves are one of the influences that shape preferences. They are not, however, identical and, as stated previously, neither is reducible to the other. A political culture is not simply the sum of individual preferences, nor do preferences, especially those of any given individual, necessarily correspond with normative prescriptions. Indeed, the mismatch is at times quite glaring. Yet it is essential to understand the goals that are embodied in both individual preferences and normative prescriptions, for it is the dynamic relationship between the two that underwrites the stability or instability of political systems.

Each of the political culture approaches has strengths and weaknesses with regard to the relationship between political culture and preference orientations. In the *hermeneutic approach*, for example, careful fieldwork involving immersion in local networks reveals the myths, symbols, and rituals that convey the meaning and intensity of the norms that underwrite social solidarity. There is a careful and nuanced description of sets of interrelated normative prescriptions. The researcher delineates which symbols undergird authority, how these symbols came into being, and how their salience is made known through a variety of rituals. The analysis is rich in detail but is concomitantly highly descriptive. Why people choose to believe in certain symbols or are swayed by particular rituals is generally inferred from analysis of social action and discourse. In explaining change the presence of opposing preference orientations is usually adduced by observation of an alternative, or counterhegemonic, discourse that references an "unoffi-

cial" set of myths, symbols, and rituals. Interpretive analysis of behavior is privileged and surveys and/or psychological interviews tend to be avoided.

Kertzer makes the case succinctly when he speaks of political change as encapsulated in a narrative about the party and its history. Occhetto, the leader of the Italian Communist Party, proposed "that the narrative itself change and, with it, the identity and history of those who based their personal narratives on it" (p. 155). Narrative in this sense is a proxy for preference orientations. These dispositions are then said to be connected to symbolic associations and forged through ritual, which fosters "heightened emotional states" that link the cognitive aspects of political allegiance and political activism "both with particular social bonds and with a particular world view" (p. 127). Notwithstanding the intent of this formulation, the cognitive aspect is largely asserted. Kubik, in his excellent study of legitimacy in Poland, states clearly where the focus of analysis lies: "The introduction of sociological surveys, with their specific methodology and results, would destroy the thematic, theoretical, and methodological unity of the book, whose primary concern is with *the interpretation of symbolic behavior*."[16]

Culture theorists adhere to the opposite extreme. In their concern to avoid ad hoc description, they embrace a theoretical model in which only a stipulated number of ideal-type preference orientations can exist. This conception is necessarily coercive and arbitrary, as macropolitical culture (the level familiar to most people) virtually drops from sight. Instead, political culture is interpreted as the preference orientations of individuals.

A strength of culture theory is its parsimony in mandating the recognition of five (no more and no less) different preference orientations. It requires, as well, an understanding of interactions among people with different preferences. The preferences themselves, however, are said to be derived from conscious choices that individuals make regarding an appropriate lifestyle. This ignores important socialization influences and unconscious motivations (for example, aggressive impulses) that may also affect preference orientations.

Culture theory incorporates an explanation for why an individual's preference orientation may shift—because of dissonance caused by an accumulated mismatch between the promise and the performance of an individual's chosen lifestyle. However, because political culture is conceived of as microculture, that is, the preferences of individuals, there

[16] Kubik (fn. 5), 10, emphasis in original.

has been less exploration of dominant configurations at the macrolevel. A general typology of such distributions for macropolitical cultures has not been attempted. (A start has been made, however, for the United States, which is said to have two dominant orientations, egalitarianism and individualism, and for Britain, where the dominant orientations are said to be hierarchalism and egalitarianism). The relationship between micropolitical culture (the preferences) and macropolitical culture is thus poorly articulated. There is no convincing explanation of how dominant configurations at the macrolevel influence the choices of individuals. The problem of feigned compliance, where people express particular preferences in order to avoid punishment, is also unexplored.

The *social character approach* recognizes the broad differences that exist among political cultures, employing criteria such as language or religion to assess significant differences from society to society. Although this approach is ideally suited for generalizations about large-scale differences among political cultures, these sweeping descriptions impose a uniformity on society's members that does not exist. The consequence is that political cultures are reduced to little more than a description of dominant cultural themes.

Preferences in the social character approach are presented as group preferences, derived from a shared socialization experience, with the consequence that they are virtually synonymous with cultural norms. The dynamic interaction between political-cultural norms and individual preferences is elided. Indeed, they merge. As Richard Solomon put it regarding the Chinese: "The notion of a dependency social orientation as the modal personality configuration . . . establishes conceptual linkages between the system of personality and the larger cultural tradition."[17] That there may be a significant number of people within the group who hold alternative preference orientations is often not acknowledged, nor does analysis of political transformation generally proceed by reference to the dynamic interaction among people who hold different preference orientations.

The *social learning approach* provides an exhaustively detailed account of dominant and emergent preference orientations within societies. Based on an assumption of shifts in socialization practices related to changes in variables at the social level (for example, increasing affluence), the approach specifies the relative salience of modal preference orientations at any given point in time. However, in Inglehart's work

[17] Solomon, *Mao's Revolution and the Chinese Political Culture* (Berkeley and Los Angeles: University of California Press, 1971), 5.

the approach places preferences in a dependent relationship to cultural, political, and economic variables. Two dimensions, survival/well-being and traditional authority/secular rational, define a field; differences among societies in their materialist-postmaterialist value orientations are mapped on this field according to their position relative to the pole values of the two dimensions. Only the values associated with one dimension (traditional authority/secular rational authority), however, could be construed as embodying political-cultural norms, and that dimension has meaning as an explanatory variable only in conjunction with economic and social variables. Hence, the cultural merges with the social/economic and is not treated as having independent validity.

If culture entails normative prescriptions that result from the interaction of sociological and psychological variables, but is not reducible to either, then Inglehart needs to focus more attention on culture per se if he truly wishes to retain it as an independent variable. Otherwise, culture is not clearly distinguished from the value orientations (materialism and postmaterialism) that are presumed to be the product of cultural (and economic and political) influences. As his work clearly reveals, the weight of the socialization hypothesis falls instead largely on political and economic variables.

The *cognitive epistemological approach* sees societies as composed of people who reason differently depending on their stage of development. Societies thus always contain individuals with different preference orientations. As a consequence, the normative component of moral and ethical norms, the political culture, is subject to continuous reassessment. The approach has been criticized for its conception of change (although without reference to the theory itself).[18] It does provide an explanation for transformations in political cultures, in part as a consequence of the emergence of new modal preference orientations that are infused with developmentally higher levels of moral reasoning.

In this approach political cultures and preference orientations are distinct areas of analysis. The dynamic forces that foster changes at both levels are explained by reference to Piagetian theory. The resolution of "dilemmatic concerns" is a major conceptual device for understanding why individual preference orientations change and why there are shifts in the moral and ethical norms of political cultures.[19] Such concerns often address the functional requirement of political cultures both to legitimize status inequalities and to underwrite social solidarity.

[18] David D. Laitin, "The Civic Culture at Thirty," *American Political Science Review* 89, no. 1 (1995).

[19] Michael Billig et al., *Ideological Dilemmas: A Social Psychology of Everyday Thinking* (London: Sage Publications, 1988), 18, 25.

But a significant body of empirical data is still lacking. Thus more clearly needs to be done to clarify (and verify) the linkage between the two levels of analysis.

There is, then, a general failure of most approaches to recognize the dynamic interaction between political culture and preferences. Serious implications flow from this: a theory of political culture that slights preferences risks being little more than a description of cultural norms; a theory that slights the normative dimension in favor of an analysis of preferences risks being no more than a typology of individual values. A fully formed theory, therefore, must take account of the fact that political cultures are not simply the product of preferences and that preferences are not simply a reflection of the political culture. They are, rather, inextricably interdependent in a manner that requires theoretical explanation at both levels and of the linkage between them.

LINKAGE

Linking the norms of a political culture with preference orientations is the most difficult theoretical task faced by political culture theorists. As linkage implies a reciprocal relationship, a theory of linkage must stipulate what changes at either the cultural or the personality level will stimulate change at the other level and what mechanisms foster interaction. It must explain how preferences act upon culture and, conversely, how normative prescriptions affect individual choices. Thus, for example, if massive and rapid technological change alters the moral justifications that legitimize status, then prevailing normative prescriptions, if unchanged, will become less effective as coordinators of social behavior and as an influence on preference formation. If the introduction of novel ideas alters individual preferences, then normative prescriptions, if unchanged, may become incongruent with the moral preferences of an increasing number of people. Changes in meanings at either level are thus interactive; they are related to the mixing of people from different cultures, the invention or diffusion of new relationship patterns, changes in social complexity, technological or environmental changes, the introduction of new ideas, and the capacity of individuals to assess and compare moral alternatives.[20]

Linkage thus requires a correspondence of meanings; normative prescriptions must be comprehensible (if not universally legitimate) to the

[20] Alan Page Fiske and Philip E. Tetlock, "Taboo Trade-offs: Reactions to Transactions That Transgress the Spheres of Justice," *Political Psychology* 18, no. 2 (1997), 284.

members of society, while individual preferences must resonate with selected aspects of normative prescriptions. In addition, there must be some way to transmit meanings between levels. I suggest that transmission has two aspects: (1) communications about political-cultural norms to the members of society (for example, socialization, ideological campaigns, party rallies, and legal decisions) and (2) communications among individuals that evaluate political cultural norms (for example, moral discourse, activism, dissident literature, and voting). Evaluating linkage for degrees of congruence in meanings reveals important information about the stability of political systems and clues about which groups are central or marginal in the political process.

The question of correspondence of meaning, however, is thorny given the widely different assumptions and methodologies employed in the various approaches. I contend, therefore, that the normative prescriptions of political cultures must be viewed in terms of how public morality is addressed—how right and wrong are stipulated in social life, who gets what kind of reward and in what amount, how particular property rights regimes are legitimized, and so on. Given this feature of political cultures, it follows that the most fundamental axis around which communications of meaning revolve is moral in nature, linking in some fashion the ethical standards embodied in normative prescriptions with the moral sentiments of individuals.

How then might theorizing about moral linkage enrich studies of political culture? Four avenues for future research will be noted. All emphasize morality but do so in different ways. The first refines the concept of preference formation to include an explicit moral dimension. The next two focus on linkage itself. The last presents a startling new vision of culture with linkage defined in biological terms. Each requires further elaboration and testing and ultimately greater scrutiny in the marketplace of ideas.

First, while many works on political culture use a combination of approaches[21]—placing variable emphasis on surveys, historical analysis (that is, path dependency), and understanding the themes embedded in myths and rituals—they rarely, if ever, attempt to evaluate the merits of the approaches or develop theory that links preferences and political culture. An effort in this regard has obvious value. For example, it is clear that the norms of cultures are learned and must be widely shared for social viability. At the macrolevel shared values, ascertained in the social learning approach through surveys, are the general wiring dia-

[21] Putnam (fn. 13).

gram that knits the members of society together. Because the process of uncovering these values is subject to statistical controls, the determination of patterns is less subject to error than when they are decoded by expert identification. Survey research by itself, however, is insufficient for understanding changes in cultural norms. Longitudinal studies to ascertain change, based on the assumption of modifications in the learning environment, say nothing about why in fact people change their values and beliefs.

To understand the dynamic of change it is necessary to determine the relative strength of the social relationship patterns that people choose to be identified with, the moral bases for these choices, and the way that these choices relate to the moral quality of political-cultural norms. This takes analysis into the realms of culture theory and cognitive epistemology. Culture theory posits that change occurs because individuals experience a cumulative disjunction between values and opportunities and hence shift their choice of social relationship pattern, leading ultimately to differences at the macrolevel in the relative strength of the patterns. Cognitive epistemological theory, however, posits that individuals develop morally with age, and thus as they mature apply different standards of evaluation to cultural rules. As individuals change, they will choose social relationship patterns that more closely accord with their changed moral competencies. Adding this explanation of preference change to culture theory's notion of "surprise" thus strengthens the underpinning of interpretations of the changes that can occur in preference orientations. Overall shifts in the relative salience of different preference orientations will then provide clues for reciprocal changes in normative prescriptions.

A second avenue for exploration focuses on understanding the mechanisms that link how individuals comprehend moral norms and how these normative injunctions are formed and made known. One place to start is by reference to North's definition of institutions (the major constraints on social relations) as "a set of rules, compliance procedures, and moral and ethical behavioral norms designed to constrain the behavior of individuals in the interests of maximizing the wealth or utility of principals."[22] On the basis of this formulation, political culture can be conceptualized as the "moral and ethical behavioral norms" that undergird rules (noted preeminently in the legal culture) and that stipulate the ways that rules are enforced (that is, as "compliance ideolo-

[22] Douglass C. North, *Structure and Change in Economic History* (New York: W. W. Norton, 1981), 201–2.

gies").[23] Linkage is thus the interactive relationship between the moral prescriptions of the political culture and the moral preferences of individuals. Specifically, there is a reciprocal relationship (that is, a correspondence of meaning) between moral constructions of meaning at the political-cultural level regarding rights and duties and moral constructions of meaning at the individual level regarding expectations and social obligations.

In this conception the moral rules that undergird political-cultural norms are expressed in terms of culturally specific definitions of rights and obligations. Political cultures, both contemporaneously and historically, reflect different ways that rights and obligations are defined and honored. These result in different definitions of deviance, different ways that political legitimacy is secured, differences in who may participate in the political process, different limitations on the exercise of power, differences in the type and extent of the rewards that elites may legitimately claim, and variations in likely trajectories of change. Although the differences among societies may be large, the focus on rights and obligations makes it possible to compare different political cultures.

At the individual level an examination of moral discourse (how and when people talk to each other about rights and obligations) aids in understanding why status rules change (as, for instance, changes in the role of women that have occurred in many societies over the past hundred years). Changes such as this occur in response to a variety of influences (such as technological change, famine, and defeat in war) but always involve a moral questioning of prevailing normative standards.[24] The result is by no means always the swift alteration of cultural rules, for powerful countervailing forces may block the development and diffusion of more inclusive moral standards. What does occur is a debate, often fraught with danger, about the possibilities inherent in a transformed political culture.

A third promising avenue for further research is exemplified in the work of Michael Gross. Gross grounds preference orientations in two theoretical frameworks, moral development theory, and rational choice theory.[25] He argues that the development and maintenance of a demo-

[23] Wilson (fn. 15).

[24] Richard W. Wilson, "American Political Culture in Comparative Perspective," *Political Psychology* 18, no. 2 (1997).

[25] Another work that utilizes a rational choice perspective although without specific reference to political culture or moral prefeences is James D. Fearon and David D. Laitin, "Explaining Interethnic Cooperation," *American Political Science Review* 90, no. 4 (1996).

cratic political culture cannot be based on moral preferences alone. Rather, individuals make rational calculations in small group contexts that further the group's ability to pursue moral objectives. This formulation locates political culture within a complex sociological field that includes the organizations within which people strive for the realization of particular goals. Preferences and political cultures do not interact directly but through intermediate groupings whose internal structures are characterized by solidarity pressures, face-to-face relations, and immediate hierarchical concerns. These are the contexts within which people actually live and where, to get along, they make prudential calculations as well as moral judgments. It is an arena where preferences have an immediacy and a relevance that is lacking when political culture is conceptualized only at the societal level. The consequence is an enriched understanding of the linkage between individual preferences and political cultures. A clear theoretical advantage is that rational choice is dealt with on its own terms (as an aspect of incentive theory and organizational theory), rather than being reconceptualized as utility calculations bounded by moral preferences.

Finally, work in evolutionary psychology (or, more broadly, universal Darwinism) places cultural studies in the controversial framework of sociobiology. Specifically, it is suggested that biological evolution, rooted in the way that genes are copied and spread, has an analogue at the cultural level in memes. These are ideas, beliefs, instructional codes, and norms that are passed from person to person by a process of imitation. Organic evolution and cultural evolution are said to resemble each other in the sense that both genes and memes are replicators that obey the general principles of evolutionary theory, carrying out the crucial evolutionary tasks of mutation, replication, and selection. Both are associated with the concept of reciprocal altruism, which functions (albeit weakly) as a preference orientation. At the genetic level altruism favors animals who reciprocate friendship (thus allowing their genes to be copied at a rate greater than for those who do not reciprocate); at the level of memes altruism is rooted in moral sentiments and ideas of justice.

To date no theory of political culture has been elaborated, although game theory has modeled memetic strategies favoring cooperation that are exceptionally robust (for example, Tit-for-Tat). What is now needed is a tighter specification of the relationship of altruism to preferences, of how other preference orientations affect altruism, of memes as replicators of altruistic norms, and of the link between them. Rapid advances in the life sciences in understanding the genetic bases of be-

havior suggest that such theories will not be long in coming to the social sciences, almost certainly within the next decade.[26]

CONCLUSION

A theory of political culture must do four things. It must specify the functional role of political culture and define the concept in a way that maintains an analytical distinction between social and psychological variables. It must ground preference formation in a theory of motivation that specifies how and why people make choices that affect political life. It must specify the nature of the interaction between the normative level and the level of preference formation and show how particular interconnections (for example, socialization, moral discourse, and activism) link the two levels and how and why behavior or discourse generated at one level has an affect on the other. Finally, it must specify a sociological context in which hypotheses can be fruitfully tested.

The debate about political culture is an old one, yet at the same time it is in its infancy. New conceptualizations are challenging received wisdom in ways that enrich the culturalist-rationalist dichotomy set forth by Eckstein a decade ago.[27] The field is now poised to move beyond the sterile description of effects to an understanding of basic causes. From the vigorous debate now taking place there has arisen the possibility of a new and invigorated theory of political culture.

[26] See, as examples, Richard D. Alexander, *The Biology of Moral Systems* (New York: Aldine De Gruyter, 1987); Robert Axelrod, *The Evolution of Cooperation* (New York: Basic Books, 1984); Jerome H. Barkow, Leda Comides, and John Tooby, eds., *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (New York and Oxford: Oxford University Press, 1992); Susan Blackmore, *The Meme Machine* (New York: Oxford University Press, 1999); Richard Dawkins, *The Selfish Gene* (New York and Oxford: Oxford University Press, 1976); William H. Durham, *Coevolution: Genes, Culture and Human Diversity* (Stanford, Calif.: Stanford University Press, 1991); and Robert Wright, *The Moral Animal: Evolutionary Psychology and Everyday Life* (New York: Vintage Books, 1995).

[27] Harry Eckstein, "A Culturalist Theory of Political Change," *American Political Science Review* 82, no. 3 (1988).

STATEMENT OF OWNERSHIP, MANAGEMENT, AND CIRCULATION, 9/27/99

Title: *World Politics*. Pub. No. 0043-8871. Frequency: Quarterly. Four issues published annually. Subscription price: $86.00 institutions, $28.00 individuals. Location of office of publication. The Johns Hopkins University Press, 2715 North Charles Street, Baltimore, MD 21218. Headquarters of publishers: Same. Publisher: The Johns Hopkins University Press, 2715 North Charles Street, Baltimore, MD 21218. Editor: Ilene Cohen, World Politics, Bendheim Hall, Princeton, NJ 08544. Owner: Princeton University, Princeton, NJ 08544. The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes have not changed during the preceding 12 months.

Extent and nature of circulation		Av. no. copies each issue preceding 12 mos.	Actual no. copies single issue pub. nearest to filing date
A	Total no. copies printed	3728	3745
B.	Paid circulation, mail subscriptions	3206	3337
C.	Total paid circulation	3217	3348
F.	Free distribution	134	137
G.	Total distribution	3351	3485
H.	Copies not distributed	377	260
I.	Total	3728	3745

I certify that the statements made by me above are correct and complete. William M. Breichner, Production Manager.

ABSTRACTS

IDENTITY AND PERSUASION

HOW NATIONS REMEMBER THEIR PASTS AND MAKE THEIR FUTURES

By CONSUELO CRUZ

Identity struggles are once again a salient problem in world politics. This article aims to throw light on the sources, dynamics, and consequences of identity formation and mobilization. It makes two theoretical arguments. First, because collective memory is both a seemingly factual narrative and a normative assessment of the past, it shapes a group's intersubjective conceptions of strategic feasibility and political legitimacy. This is why collective identity is above all an expression of normative realism: a group's declaration to itself and to others about what it can or cannot do; what it will or will not do. Second, at critical junctures competing actors assert or contest the normative realism underlying collective identity. They do this through rhetorical politics, deploying their powers of persuasion in order to engage the constitutive elements of the group's shared identity. In practical terms, rhetorical politics is structured by a dominant frame: a historically shaped discursive formation that does two things. It articulates in readily accessible ways the fundamental notions a group holds about itself in the world and allows or disallows specific strategies of persuasion on the basis of their presumptive realism and normative sway. Within this frame, rhetorical politics engenders a *collective field of imaginable possibilities:* a restricted array of plausible scenarios about how the world can or cannot be changed and how the future ought to look. Though circumscribed, this field is vulnerable to endogenous shifts, precisely because actors' rhetorical struggles introduce conflicts over the descriptive and prescriptive limits of what is "realistically" possible. Such conflicts may in fact produce a new dominant rhetorical frame and profoundly influence a nation's political and economic development. Two contrasting cases from Latin America offer empirical support for these arguments. The article shows that the sharp developmental divergence between Costa Rica and Nicaragua can be properly understood only through close analytical scrutiny of the different rhetorical frames, fields of imaginable possibilities, and collective identities that rose to prominence at critical points in these countries' colonial and postcolonial histories.

THE CAUSES OF WELFARE STATE EXPANSION

DEINDUSTRIALIZATION OR GLOBALIZATION?

By TORBEN IVERSEN and THOMAS R. CUSACK

An influential line of argument holds that globalization causes economic uncertainty and spurs popular demands for compensatory welfare state spending. This article argues that the relationship between globalization and welfare state expansion is spurious and that the engine of welfare state expansion since the 1960s has been deindustrialization. Based on cross-sectional-time-series data for fifteen OECD countries, the authors show that there is no relationship between globalization and the level of labor-market risks (in terms of employment and wages), whereas the uncertainty and dislocations caused by deindustrialization have spurred electoral demands for welfare state compensation and risk sharing. Yet, while differential rates of deindustrialization explain differences in the overall size of the welfare state, its particular character—in terms of the share of direct government provision and the equality of transfer payments—is shaped by government partisanship. The argument has implications for the study and the future of the welfare state that are very different from those suggested in the globalization literature.

WAGE INEQUALITY AND VARIETIES OF CAPITALISM

By DAVID RUEDA and JONAS PONTUSSON

This article draws on a new data set that enables the authors to compare the distribution of income from employment across OECD countries. Specifically, the article conducts a pooled cross-sectional time-series analysis of the determinants of wage inequality in sixteen countries from 1973 to 1995. The analysis shows that varieties of capitalism matter. The authors find that the

qualities that distinguish social market economies from liberal market economies shape the way political and institutional variables influence wage inequality. Of particular interest to political scientists is the finding that the wage-distributive effects of government partisanship are contingent on institutional context. Union density emerges in the analysis as the single most important factor influencing wage inequality in both institutional contexts.

REASSESSING THE THREE WAVES OF DEMOCRATIZATION

By RENSKE DOORENSPLEET

Since the publication of Samuel Huntington's 1991 study of democratization, scholars have come to take for granted the notion that the spread of democracy has come in waves. Although Huntington's work has clearly proved to be an influential study, this article suggests that his analysis is far from compelling. There are two problems embodied in the work. The first is largely conceptual. Huntington's analysis fails to provide a clear and meaningful distinction between democratic and authoritarian regimes because it focuses primarily on what Dahl had defined in 1971 as the dimension of competition and pays insufficient attention to the equally important dimension of inclusion. The second problem is more empirical. Huntington has estimated the incidence of transitions to democracy in terms of the *percentages* of world states involved. Since the denominator in this equation, that is, the number of states in the world, is far from constant, this measure can be misleading. This article proposes solutions to both of these problems, and this new approach leads to conclusions that are quite different from Huntington's.

The new findings are important not only because they suggest a somewhat altered sequencing and a more accurate count of democratization waves but also because they cast real doubt on the appropriateness of the wave metaphor. As the reverse waves are not really apparent from these data, these results therefore also indicate that future studies should be cautious in comparing, explaining, and forecasting different "waves" of democratization.

THE STUDY OF MONEY

By JONATHAN KIRSHNER

Monetary phenomena define the contours of the contemporary global economy. This is a recent development, and it will transform the study of international political economy (IPE). Two excellent new books, *The Geography of Money*, by Benjamin Cohen, and *Mad Money*, by Susan Strange, will frame, support, and provide the point of departure for scholars addressing this vital question. Ultimately, however, and perhaps necessarily, these books raise more questions than they answer. But they do suggest in which direction the most promising avenues of investigation point—toward the study of the unique interconnections between the ideas, material interests, and institutions associated with the management of money. Those relationships are profoundly consequential for politics and demand the renewed attention of contemporary scholars of international relations and political economy.

IDENTITY AND PERSUASION
How Nations Remember Their Pasts and Make Their Futures

By CONSUELO CRUZ*

IDENTITY politics now occupies center stage. It is on explosive display in a wide range of settings, from Africa to the Balkans. And it is more subtly at work in the democracies of the West: NATO leaders must reexamine their own guiding principles in the face of multiple humanitarian disasters brought about by tribal, ethnic, and religious hatreds, while simultaneously dealing with increasingly heterogenous populations at home and ceaseless migrant flows from east and south.

It is no surprise, then, that identity should figure prominently in scholarly debates about conflict resolution, migration, citizenship, transnational alliances, democratization, nation and state building, and even globalization. Nor is it a puzzle that long-standing debates about identity should have produced the two clearly defined camps that we know as primordialists and constructivists.[1] What is perplexing is the current failure to investigate systematically the complex linkages between identity formation and the expressive practices of political actors. After all, we approach reality rhetorically—that is, with an intent to convince.[2] This is why persuasion is the hallmark of effective leadership. Short of pure coercion, it is the most direct way to mobilize or paralyze a group. And this is why rhetorical power—recognized as cen-

* I would like to thank Peter Hall, Jeffrey Legro, Sheri Berman, and Mehmet Tabak for their insightful suggestions on a previous version of this article. I am also grateful to Deborah Yashar, Anna Seleny, Paola Cesarini, and the anonymous reviewers for their helpful comments on the present version. Any errors that remain are my responsibility alone.

[1] A third approach, instrumentalism, concentrates less on identity formation and more on the rational incentives that motivate political entrepreneurs to mobilize their "primordial" constituencies. In this article the role of political leaders is also emphasized (although I explore the popular dimension of identity formation elsewhere, in Cruz, "World Making in the Tropics: The Politics of Fate and Possibility" [Book manuscript]). But leaders' mobilizational motivations and capabilities, as well as rationality itself, are treated here as embedded in the broader political-cultural conditions shared by their respective groups. A more detailed discussion on this point follows in the text below, with reference to David Laitin's work.

[2] J. Gusfield, *Kenneth Burke: On Symbol and Reality* (Chicago: University of Chicago Press, 1989), 13, 29.

tral to the workings of the polity since classical antiquity—is deployed by the full range of political agents, from aboriginal chiefs in recondite regions of South America to professional politicians on Capitol Hill.[3]

This article makes two arguments. The first is that we cannot grasp the nature and dynamics of political identity—and collective identity more broadly—unless we understand the rhetorical frames that emerge as dominant at critical junctures in the history of a group or a nation. Indeed, we will see that it is precisely because identity, memory formation, and rhetorical frames are so closely entwined that identity has both the appearance of naturalness emphasized by primordialists and the constructed properties emphasized by constructivists. The logic here is as follows. Collective memory, by its very nature, impels actors to define themselves intersubjectively. Shaped by past struggles and shared historical accidents, collective memory is both a common discriminating experience (this was right, that was wrong) and a "factual" recollection—a seemingly veridical narrative—of the group's past "as it really was."[4] Thus whether in war or in peace, a collectivity expresses and defends its identity by declaring, "We are as we are because the world has made us this way; and because we are who we are, we can change our world only so much without changing ourselves."

This declarative imperative does not suggest, however, that identity is reducible to a rhetorical manifesto. Rather, it simply suggests that while the impetus for identity formation arises from collective memory and contextual changes, identity is ultimately contingent on a "realistic" description of the world and on a relatively strict understanding of the permissible and the forbidden. And on neither of these counts can collective self-definition be completed without the active participation of those living the identity. At critical points, for example, rival political leaders and entrepreneurs seek to persuade themselves and others that things must either remain as they are or be changed in significant ways. In their efforts at persuasion, rivals must appeal to the national group's sense of practical competence (its sense of mastery over the

[3] For a superb exploration of the role of eloquence in tribal politics, see Pierre Clastres, *Society against the State* (New York: Zone Books, 1989).

[4] This view of collective memory hinges on the narrative construction of history while at the same time eschewing the presumed split between the normative and cognitive thrusts behind narrative construction. For some, like Hayden White, the objective behind the construction of sequenced history is to impose a normative claim to authority. For others, like Louis O. Mink, narrative construction is a cognitive enterprise. See White, *Metahistory* (Baltimore: Johns Hopkins University Press, 1973); idem, "The Value of Narrativity in the Representation of Reality," in W. J. T. Mitchell, ed., *On Narrative* (Chicago: University of Chicago Press, 1981); and Mink, "Everyman His or Her Own Annalist," in Mitchell. For a succinct discussion of the split, see David Carr, *Time, Narrative and History* (Bloomington: Indiana University Press, 1986).

world) and to its convictions about the possible (what the group can or cannot do; and normatively, what it must or must not do). In brief, rivals seek a political grip on the constitutive elements of collective identity.[5] This they cannot do at will, however. They must advance their competing visions and agendas within a dominant rhetorical frame—a discursive structure that articulates in accessible ways the fundamental notions a group holds intersubjectively about itself in the world and that allows or disallows specific strategies of persuasion on the basis of their presumptive realism and normative sway.

The second argument this article makes is that the (trans)formation of collective identity shapes a nation's political and economic development.[6] The logic of this second argument is closely related to that of the first. Because actors situate their struggles within a dominant rhetorical frame, political contests between them engender a *collective field of imaginable possibilities*, which I define as a restricted array of plausible scenarios of how the world can or cannot be changed and how the future ought to look. This field's boundaries are established at critical points, along with new sets of power relations and the rhetorical settlements that accompany their construction. Within such boundaries, actors routinely make claims to vocality, manipulate the positive and negative values assigned to past defining experiences, generalize from these claims and experiences to craft "simply is" statements about reality, and drawing on this generalization, identify viable routes—be they conservative or transformative—to a better future.

But since a collective field of imaginable possibilities is both a system of meanings (attached to vocality, past experiences, reality, and the future) *and* an arena for the play of rhetorical practices, it is also internally vulnerable to endogenous shifts. Indeed, actors' rhetorical struggles can introduce disorientation into the field and, by extension, challenge the dominance of a particular rhetorical frame.[7] Here the microdynamic is simple. Competing political leaders and entrepreneurs

[5] This claim is congruent with the notion that national-identity formation is inseparable from the interplay of actors' other political identities and agendas. See Prasenjit Duara, "Historicizing National Identity, or Who Imagines What and When," in Geoff Eley and Ronald Grigor Suny, eds., *Becoming National* (New York: Oxford University Press, 1996).

[6] Development is not conceived here as a fixed end point. Rather, development is the ongoing, effective creation of institutional conditions that (1) a critical mass of agents deems necessary for the rational pursuit of economic prosperity and that (2) allow for the establishment of a viable system of political legitimation, thus making possible the relatively smooth distribution of material resources and the stable allocation of authoritative functions and prerogatives among rulers, officials, and citizens.

[7] Even in the natural sciences, individuals live with doubt, to which they respond in a variety of ways, ranging from isolated tinkering to heated debate and paradigmatic discovery. Thomas S. Kuhn, *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press, 1970).

intuit Edmund Burke's dictum that "it is one thing to make an idea clear and another to make it affecting to the imagination."[8] And so they seek at least one of two goals. Political leaders and entrepreneurs aim to (re)cast a particular agenda as most appropriate to a given collective reality or to recast reality itself by establishing a (new) credible balance between the known and the unknown. In the former instance they aim to establish the facts and weave them into a whole that is likely to be accepted as both "realistic" and "just." In the latter, they aim to reshape the collectivity's intersubjective "realism" and its sense of viable justice—in short, to redefine the limits of the possible, both descriptively and prescriptively. At such moments collective identity itself is open to revision, actors' choices and power relations may be reconfigured, and leaders may either gain or lose the ability to craft and execute developmental programs.[9]

The empirical evidence offered for this argument is a comparative study of the role of collective identity in molding the postcolonial development of two contrasting cases: Costa Rica and Nicaragua. The comparative starting point is the formation of colonial power relations, collective memories, and rhetorical legacies. Once these are established, the analytical focus shifts to the two distinct types of strategic rhetorical practices that rose to prominence in the two countries in the wake of independence. We might call one type the "Alchemist's Trick"; this was the Costa Ricans' way. Another was the Nicaraguan option; call it the "Fatalist's License." Through the former, rivals minimized the threat they posed to each other by emphasizing their shared virtues. Through the latter, they sought justification for bending the rules to selfish advantage by stressing their common flaws.

The more postcolonial leaders deployed these strategies, the more they entrenched (often unwittingly) two contrasting but equally powerful rhetorical frames. The Costa Ricans' frame upheld as truth the claim that they were, both by reputation and in fact, an inherently peaceable and diligent people. This frame, in turn, restricted the intensity and scope of political debate and struggle, creating in the process sufficient space for intraelite consensus building and, ultimately, the political stability necessary to craft and execute developmental policies. In the case of postcolonial Nicaragua, the argument is exactly the reverse.

[8] Cited in Virginia Sapiro, *A Vindication of Political Virtue* (Chicago: University of Chicago Press, 1992), 190–91.

[9] Thus the dichotomy between culture as a system of meaning and culture as practice dissolves in a collective field of imaginable possibilities. For an excellent critique of this dichotomization, see William H. Sewell, Jr., "The Concept(s) of Culture," in Victoria Bonnell and Lynn Hunt, eds., *Beyond the Cultural Turn* (Berkeley: University of California Press, 1999), 46–47.

The rhetorical frame that triumphed there generated a set of negative expectations that created incentives among political rivals for *preemptive, particularist* bids, thus closing off space for effective developmental leadership. The upshot of all this was the emergence of distinct national identities and equally distinct collective fields of imaginable possibilities—one as "civic" and optimistic as the other was "anarchic" and fatalistic.

I. Positioning the Argument

Primordialists take identity as fixed by inherited linguistic, racial, ethnic, and/or territorial commonalities. For constructivists, however, identity is tied not to a marker but to the civic and political commonalities that bind a national group together. But not even these commonalities are fixed, since national tradition itself is a political invention. Moreover, for constructivists, identity can be contingent on choices that individuals make within shifting sociopolitical and historical environments, especially in response to the bids of entrepreneurs who trade in the business of identity politics.[10]

The view of identity advocated above—call it an argument for *declarative identity*—differs sharply from primordialist definitions. Primordial markers matter in cases where linguistic, religious, and racial groups are in conflict or where exclusivist claims to a particular territory are at play.[11] But such markers are less useful in getting at the core of collective identity in relatively homogenous settings where they are not used by the actors themselves or where these markers have been either marginalized in the public arena or subsumed under a broader worldview. Constructivist accounts, by contrast, tend to be flawed in one of two ways. In the civic and political variant, identity is often treated as a functional extension of modernization and nation building.[12] The invention-of-tradition variant, in turn, risks voluntarism by stressing the will and capacity of nation builders in the invention and sacralization of collective identity.[13]

[10] Clifford Geertz, *The Interpretation of Cultures* (New York: Basic Books, 1973).

[11] For an example of the argument that ethnicity is the core determinant of national identity, see Walker Connor, *Ethnonationalism: The Quest for Understanding* (Princeton: Princeton University Press, 1994). And for the claim that the ethnic factor is always present to some degree in the making of national identity, see Anthony Smith, *National Identity* (London: Penguin Books, 1991).

[12] Ernest Gellner, for one, treats national identity as necessary to modernization and nation building. See Gellner, *Nations and Nationalism* (Oxford: Basil Blackwell, 1983).

[13] See Eric Hobsbawm and Terence Ranger, eds., *The Invention of Tradition* (Cambridge: Cambridge University Press, 1983).

My argument for declarative identity builds selectively on constructivist scholarship by stressing the intersubjective, constructed properties of identity, but it tries to avoid functionalism and voluntarism in several ways. First, I start from the premise that invention is a constrained activity. Consider here the credibility dilemma we all confront at one point or another. On the one hand, we cannot always remember everything. On the other, when faced with a momentous choice, we need to feel as if we do remember it all, and so we invent what we cannot remember. Yet a historical invention must be *credible;* it must ring true. Political leaders and entrepreneurs understand this well. They also know that the ring of truth is often no more than the ring of the familiar. Thus at critical moments political rivals will draw on the same rhetorical frame—the same fundamental, easily graspable "truths"—to advance their contending views of past and future. Indeed, this tendency to turn to easily graspable truths is partly why, as Radman reminds us, "invention is little more than a new combination of those images which have been previously gathered and deposited in the memory."[14] (In the late eighteenth century, for example, North American colonials were able to articulate simultaneously their allegiance to Great Britain and their differences with the British monarchy by drawing on images embedded in English legal and political tradition).[15]

The argument for declarative identity, moreover, admits that while political invention is a constrained activity, invention may also happen inadvertently. The underlying claim here is that rhetorical politics can destabilize actors' purposeful aims. Contending arguments, for example, need not remain a linear projection of shared premises, since they interact in unpredictable ways with the behaviors and events they promote. In addition, the changing demands of competition prompt rivals to make rhetorical refinements, elaborations, and connections that in turn may cause shifts in the collective field of imaginable possibilities. (North American colonial elites also rendered the radical idea of independence both plausible and justifiable, in part because their own debates led to the creative deployment of the very same tradition of English law and political theory that had once tied them to Great Britain.)

Finally, my argument for declarative identity assumes that cultural entrepreneurs and their potential clients are embedded in the same col-

[14] Zdravko Radman, *Metaphors: Figures of the Mind* (The Netherlands: Kulwer Academic Publishers, 1997), 63–65.

[15] Tadahisa Kuroda and Erwin Levine, "The United States: Creating the Republic," in Mary Ellen Fischer, ed., *Establishing Democracies* (Boulder, Colo.: Westview Press, 1996).

lective field of imaginable possibilities. As in Ann Swidler's "tool kit" approach to culture, all actors draw on the same "repertoire" of symbols, stories, and worldviews to craft and execute their strategies. If actors persist in their political-cultural practices, it is not because they are "cultural dopes." If they quit, it is not because they are without culture. Rather, actors in each case behave as competent "users of culture."[16] This key assumption helps us get at both the naturalness and the plasticity of identity without falling prey to the problems that plague instrumentalist approaches, like the tipping model of identity formation.

In David Laitin's rendition of the tipping model, for example, cultural entrepreneurs are treated as if they are invulnerable to the equilibrium conditions under which other members of a linguistic, racial, ethnic, and/or national group experience their beliefs and practices as natural. Cultural and political elites, that is, "give meaning to the equilibrium" by providing the group with a set of "beliefs, principles, and constraints," but elites themselves appear unencumbered by "meaning" as they respond strategically to changing incentives.[17] Meanwhile, the agency of followers and clients is split as follows. On the one hand, they are sovereign decision makers when facing a dilemma. For example, when members of a peripheral culture must choose between their mother tongue and a newly legislated *national* language,[18] they weigh the costs and benefits of "switching" according to instrumental criteria, disregarding the judgment others might make about their choices. On the other hand, from the perspective of the tipping logic what matters most to a decision maker is what others are doing because this is how individuals evaluate future costs and benefits, which is why a cascade can occur at all. The upshot of all this is that the importance of meaning—and identity's presumed naturalness—are lost in the thick of actors' instrumentalism.

Swidler's approach does not split actors in these ways, but it does presume *two* models of cultural causality: one that obtains during "settled" periods and another that obtains during "unsettled" times. In settled times culture influences behavior because it provides the resources with which agents put together lines of action. In unsettled periods "explicit ideologies" rule over action, but structural opportunities for action determine the fate of competing ideologies.[19] This duality, however,

[16] Ann Swidler, "Culture in Action: Symbols and Strategies," *American Sociological Review* 51 (April 1986), 273–86.

[17] David Laitin, *Identity in Formation: The Russian-Speaking Populations in the Near Abroad* (Ithaca, N.Y.: Cornell University Press, 1998), 22.

[18] Ibid., 25–27.

[19] Swidler (fn. 16).

begs the question of transition from settled to unsettled times and its causes. In contrast, my argument for declarative identity—with its emphasis on rhetorical frames and fields of imaginable possibilities—allows for endogenous transitions. Contestation is present when political actors construct rhetorical frames in an attempt to organize action and (re)organize reality. But contestation is also present once a particular frame has been imposed, precisely because that same frame generates internal interpretative conflicts, which may eventually lead to critical historical junctures. At such points, a dominant frame becomes one among competing frames, and its fate becomes indeterminate. This is partly why times become unsettled in the first place and shifts occur in a collective field of imaginable possibilities.

This argument for declarative identity, of course, raises its own set of questions. For instance: what is the discursively explicit, what is the practically known but unstated, and what is simply embedded in the unconscious? My argument, however, is concerned exclusively with the discursively explicit, a restriction grounded in three related considerations. First, it is extremely difficult to access actors' interior worlds. Take, for instance, ideology. As Joseph Schull argues, even individuals who belong to the same ideological group can hold divergent interpretations of the relevant ideological claims. This divergence is possible—perhaps even unavoidable—because an ideology ultimately resides in people's minds.[20] Yes, ideological treatises and platforms exist in objective form. But their content is processed and transformed internally, and for this reason, therefore, the authenticity of political agents' ideological pronouncements can be called into question, both by other actors and by observers. (Even if an individual sacrifices himself on a pyre in a sincere effort to make an ideological point, rivals can always claim that self-immolation was really motivated by vainglory or, more pathetically, by an inadequate understanding of the cherished ideology.)

Second, because of the ever-present possibility of a discrepancy between ideological tenets and people's internal interpretations, it is analytically prudent to concentrate on actors' discursive productions and manipulations. James Scott, for one, argues that in power relations between oppressors and subordinates, the latter can and do feign ideological conviction—put on masks—and keep false faith with a dominant ideology.[21] Conversely, as Timur Kuran points out, subordinates may well

[20] Schull, "What Is Ideology? Theoretical Problems and Lessons from Soviet-Type Societies," *Political Studies* 40 (December 1992), 731.

[21] Scott, *Domination and the Arts of Resistance: Hidden Transcripts* (New Haven: Yale University Press, 1990), 110.

loathe an oppressor and still fail to develop "cognitive autonomy"—fail to craft an alternative ideology—"precisely because of the social conditions created by the oppressor."[22]

Actors' enduring manipulative capability and the alternative possibility that oppression saps their ideological imagination are both plausible claims. Yet neither can be asserted in a particular case without empirical investigation. In both cases the advantage of concentrating on rhetorical frames and fields of imaginable possibilities is clear. Because discourse is an intersubjective event, as Schull argues, discourse and its individual agents are relatively autonomous from each other. Discourse is relatively autonomous since its viability does not require the faith of its agents; instead, discourse requires agents "to respect certain conventions for meaningful communication." Agents, in turn, are relatively autonomous because they need not demonstrate ideological exactitude and depth to be taken seriously; instead they must demonstrate a significant degree of consistency between their respect for discursive conventions and their own actions.[23] And since discursive conventions and actors' actual respect for those conventions (as indicated by their behavior) are both *observable* phenomena, one does not need to know agents' internal processes in order to arrive at plausible findings.[24]

Third, and finally, I focus exclusively on rhetorical politics because the observable properties of discourse are even more pronounced in rhetorical productions. Rhetorical strategies, for example, are deliberately and closely aimed at persuasion and so they generate recognizable patterns of argumentation. In fact, the more distinct a particular pattern—that is, the more different it is from rival arguments and the more identifiable it is with particular actors—the greater its potential effectiveness. But also the more exposed its vulnerable points become, and therefore the greater the ability of contending actors to organize and focus their attacks.[25] This competitive dynamic can lead either to entrenchment of a dominant rhetorical frame or to varying degrees of

[22] Kuran, *Private Truths, Public Lies: The Social Consequences of Preference Falsification* (Cambridge: Harvard University Press, 1995), 175.

[23] Schull (fn. 20), 732.

[24] Ibid., 731.

[25] Albert O. Hirschman, for instance, has shown that the rhetorical struggles associated with landmark social and economic reforms in the advanced industrial democracies produced "contrasting pairs" of arguments that not only shaped the struggles themselves but, whether progressive or reactionary, could be "exposed" as "extreme statements" in "highly polarized debates." Once exposed, such statements may be reduced to the status of "limiting cases" in need of qualification, mitigation, and/or amendment. See Hirschman, *The Rhetoric of Reaction: Perversity, Futility, Jeopardy* (Cambridge: Belknap Press of Harvard University Press, 1991), 166–67.

disarray within the collective field of imaginable possibilities—and perhaps to the emergence of a new dominant frame. Regardless of the outcome, however, the relevant meanings, practices, and structures remain analytically discernible.

II. The Case for the Cases and Contending Accounts

By the dawn of the nineteenth century, according to Benedict Anderson, Spanish American "creole communities" had developed a sense of "nation-ness"—they had begun to imagine the "fraternal" bonds that would soon hold together their individual republics, from the smallest to the largest. But as I have argued elsewhere, a close reading of actors' discursive positions and political strategies shows that creole elites relied on familial metaphors in a desperate effort to contain the "anarchy" of internal fragmentation that haunted their societies on the eve of independence and even more acutely in its aftermath.[26] Nicaragua was an extreme example of anarchic implosion, whereas neighboring Costa Rica was relatively tranquil. Moreover, while postcolonial Costa Rica soon proved a political-economic developmental success, Nicaragua was a study in failure.[27] Their contrasting fortunes make the two countries excellent cases for comparative analysis of nation and state building in postcolonial Spanish America.

Comparison of the two countries can also be enlightening on a number of additional counts. First, postcolonial societies often share one key characteristic with severely fragmented societies elsewhere: political actors typically operate under conditions of impending disarray and entrenched mistrust. Second, postcolonial societies, like today's transition societies, are in close contact with the history of a prior regime—both because collective memory is still fresh and because institutional-political legacies are still vibrant. But since those memories and legacies are substantively different from those of, say, postcommunist societies, analysis of postcolonial cases can enhance our comparative perspective and thus increase our capacity to draw conclusions about epochal transitions at an even higher level of generality. Third, and finally, postcolonial actors' success or failure in taking effective control of the polity and economy depends on their ability to regulate mistrust and institutionalize legitimate systems of arbitration and distribution.

[26] See Benedict Anderson, *Imagined Communities: Reflections on the Origin and Spread of Nationalism* (New York: Verso, 1991), 7, 50, 60. I develop my counterclaim in Cruz (fn. 1).

[27] Developmental success or failure is assessed according to the criteria outlined in fn. 6.

Thus postcolonial cases can illuminate political entrepreneurs' use of collective memories and legacies for good or for ill.

Recent transitional processes, of course, were partially shaped by demonstration effects conveyed at rapid speeds unimaginable in colonial times. We need only recall the diffusion of information from Poland to Hungary to East Germany in the late 1980s. Nevertheless, the geographical proximity of Costa Rica and Nicaragua, plus their close administrative ties under the late colonial regime, make them excellent cases for a comparative study with broader implications. Even in the last decades of the eighteenth century and the early decades of the nineteenth, administrative ties (and jealousies) prompted Costa Ricans and Nicaraguans to assume a stance of mutual observation. Moreover, while news from Guatemala might not reach either Nicaragua or Costa Rica for weeks, it took only days for Nicaraguans and Costa Ricans to become aware of their respective developments.[28] Costa Rica and Nicaragua were directly exposed to each other's demonstration effects and typically entered into an intimate communicative sphere both before and after becoming exposed to broader currents. The fact that they nevertheless proved resistant in the end to one another's influence is itself part of the puzzle to be solved. (By contrast, the transmission of information between the capitals of the Southern Cone or between Mexico City and the other major power centers in Spanish America was excruciatingly slow.)

Comparing Costa Rica and Nicaragua is also particularly helpful because scholars in both countries have shown a marked tendency to generate mirror-image national-character explanations.[29] Mainstream

[28] Guatemala was the administrative seat of the Kingdom of Guatemala, which comprised all of colonial Central America. Following independence in 1821 and the collapse of the ephemeral Mexican empire in 1823, the various Central American states formed a federal republic, which by 1838 was virtually defunct.

[29] These explanations, it should be noted, bear scant analytical relation to sophisticated national-character arguments. Alex Inkeles, for example, posits that national character refers to "the mode or modes of distribution of personality variants within a given society." The national-character explanations articulated by Costa Rican and Nicaraguan scholars, in contrast, merely echo and reinforce dominant mythologies that are part and parcel of intersubjective realisms established through long-forgotten contestations. And yet, analytically unrelated as they are, both Inkeles's argument and the homegrown national-character explanations eclipse intersubjective dynamics and contestation processes. Inkeles treats national character as the *collective* expression of *individuals'* personality, and he assumes an interrelation among modal personality, institutional structures, cultural patterns, and the actions of nation-states. By taking a cumulative rather than an intersubjective perspective, Inkeles forecloses the possibility of a group dynamic independent of its members' personality makeup. By removing the element of struggle, he reduces individuals and macrostructures to unproblematic reflections of one another. But while Inkeles's reductionism helps explain the declining influence of national-character explanations, the simplifications by Costa Rican and Nicaraguan scholars hint at some of the ways in which different rhetorical frames became entrenched in our two cases. Inkeles, *National Character: A Psycho-Social Perspective* (New Brunswick, N.J.: Transaction Publishers, 1997).

Costa Rican historiography tends to accept prima facie the "conciliatory" nature of Costa Rica's conquest[30] and to view postcolonial elites as carriers of the traits and habits first exhibited by the country's exceptionally "peaceful" conquerors.[31] Conversely, Nicaraguan historiography tends to accept uncritically the specially brutal character of the country's conquest.[32] Once vested with the authority of history, these mythologies help set the parameters for new rhetorical struggles.

Yet another reason for comparing Costa Rica and Nicaragua along the vectors of collective identity and rhetorical politics is that even when these countries' own scholars cast aside national-character explanations in favor of socioeconomic or ideology-based explanations for the countries' developmental trajectories, they still reinforce the countries' mythological legacies and render the countries' histories as chronicles of foretold success and failure. Explanations that stress divergent socioeconomic conditions at the start of the postcolonial period, for example, often share a logic that runs as follows. The Spanish metropole consistently overlooked the province of Costa Rica because it was poor in Indians. This neglect led to the politico-administrative marginalization of the province and to a widening poverty. And both these conditions in turn accustómed Costa Rica's colonial elites—and their postcolonial successors—to de facto politico-administrative autonomy and to the more egalitarian relations associated with labor-scarce colonies.[33] Conversely, Nicaragua with its large Indian population, grew accustomed to the exploitative relations that characterized better endowed colonies.[34]

But in fact, the countries' endowments of indigenous labor were not radically unequal. By the mid-sixteenth century Nicaragua's conquerors

[30] Eugenio Rodríguez Vega, *Biografía de Costa Rica* (Biography of Costa Rica) (San José: Editorial Costa Rica, 1981), 19; Carlos Meléndez, "Los Poderes Conferidos al Alcalde Mayor Licenciado Juan de Cavallón," in *IV Centenario de la Entrada de Cavallón a Costa Rica, 1561–1961* (IV Centennial of Cavallón's incursion into Costa Rica, 1561–1961) (San José: Imprenta Nacional, 1961), 41; and Jorge Lines, "Integración de la Provincia de Costa Rica bajo el Reinado de Carlos V," in *IV Centenario*, 21, 25–31.

[31] Hernán Peralta, *Don José María Peralta* (San José: Editorial Hermanos Trejos, 1956), 21. See also Juan Rafael Quesada, "Educación y Democracia en Costa Rica," in Jorge Mario Salazar, *Democracia y Cultura Política en Costa Rica* (Democracy and political culture in Costa Rica) (San José: Ministerio de Cultura, Universidad de Costa Rica, 1990).

[32] See, for example, Gustavo Alemán Bolaños, *El País de los Irredentos* (A country of irredeemables) (Guatemala City: Tipografía Sánchez y de Guise, 1927); José Coronel Urtecho, *Reflexiones sobre la historia de Nicaragua*, vol. 1 (Reflections on the history of Nicaragua) (León: Editorial, 1962); and Ernesto Cardenal, *El Estrecho Dudoso* (The doubtful strait) (Mexico City and Buenos Aires: Ediciones Carlos Lohlé, 1972).

[33] The most important exponent of this view is Carlos Monge Alfaro, *Historia de Costa Rica* (San José: Hermanos Trejos, 1980).

[34] Edelberto Torres Rivas, *History and Society in Central America* (Austin: University of Texas Press, 1993).

had virtually devastated the indigenous population; and by the early decades of the nineteenth century the abundance of land merely reinforced the labor shortage. Moreover, postcolonial Costa Rica's initial distributional structure was neither remarkably egalitarian nor extraordinarily polarized; that is, it was not substantially different from the structure that obtained spontaneously in land-abundant Nicaragua. Costa Rica's labor scarcity certainly did not translate ipso facto into an egalitarian society.[35] As William Roseberry shows: "Smallholders produced and sold their coffee—and secured labor for their coffee plots—within hierarchical social fields." Hence his caution against the "temptation" to treat "smallholder regimes" as the foundation for "rural democracy."[36]

Analyses that focus on socioeconomic interests as determinants of postcolonial ideologies are similarly flawed.[37] Costa Rican exceptionalism, for example, has been explained as the felicitous outcome of an ideological struggle between the "republican" disposition of an aborning bourgeoisie and the "retrograde" impulses of the colonial aristocracy.[38] In contrast, ideological backwardness in postcolonial Nicaragua—characterized by weak or nonexistent modern ideologies—has been treated as the superficial expression of socioeconomic undifferentiation.[39] But this ideological variant of the socioeconomistic account is contradicted by scholarship demonstrating that Costa Rican elites, like all the other regional elites, "lacked clear socio-economic or ideological lines of party affiliation."[40]

[35] Carlos Meléndez, "Bosquejo para una historia social costarricense antes de la independencia," in Carmen Lila Gómez et al., eds., *Las Instituciones Costarricenses del Siglo XIX* (Costa Rican institutions in the nineteenth century) (San José: Editorial Costa Rica, 1985), 35–48.

[36] Hierarchical distinction, while not peculiar to Costa Rica, does go against its yeoman-democracy image, especially once we take into account the fact that postcolonial elites' command of credit resources and of "administrative skills and commercial contacts" led to skewed control of coffee processing and export marketing. William Roseberry et al., eds., *Coffee, Society, and Power in Latin America* (Baltimore: Johns Hopkins University Press, 1995), 20; Samuel Stone, *La Dinastía de los Conquistadores: La Crísis del Poder en la Costa Rica Contemporánea* (The conquerors' dynasty: A crisis of power in contemporary Costa Rica) (San José: EDUCA, 1975), 75–76; and Lowell Gudmundson and Héctor Lindo-Fuentes, *Central America, 1821–1871: Liberalism before Liberal Reform* (Tuscaloosa: University of Alabama Press, 1995), 46–47.

[37] For a prominent exponent, see Rodolfo Cerda Cruz, *Formación del estado en Costa Rica* (State formation in Costa Rica) (San José: Editorial Universidad de Costa Rica, 1967), 96–99, 156–59.

[38] This account also overlooks the fact that Costa Rica's aristocracy (located in the "closed" city of Cartago) *and* the progressive elites (located in "bourgeois" San José) were one and the same, both in terms of social origins and in terms of access to national power. See Stone (fn. 36) for the shared genealogy of power holders.

[39] Jaime Wheelock, *Imperialismo y dictadura: crísis de una formación social* (Imperialism and dictatorship: Crisis of social formation) (Mexico City: Siglo XXI, 1975).

[40] Lowell Gudmundson, *Costa Rica before Coffee: Society and Economy on the Eve of the Export Boom* (Baton Rouge: Louisiana State University Press, 1986), 4.

Looking beyond the works of national scholars, we find several sophisticated accounts of Central American divergent developmental paths. Two stand out. One is Jeffery Paige's argument that these paths can be traced to the patterns of elite dominance that emerged in the export coffee economies of the nineteenth century and to how elites related to the rural masses. The Costa Rican pattern, Paige shows, can be properly understood only with close analytical reference to elite narratives and stories—most notably the "white legend" that shaped relations of production in a coffee economy in which aristocratic producers became symbiotically entwined with medium and small producers in the mid-nineteenth century.[41] Paige, however, does not theorize the broader theme of exceptionalism, nor does he link the legend's viability —either structurally or semiotically—to this theme or to the colonial antecedents of national-identity formation. A second outstanding account is provided by Deborah Yashar, who argues that democratic development is determined by institutional legacies, political coalitions, and the context of reforms.[42] These factors, undoubtedly crucial, are most revealing when placed in a full historical context—a task that calls for scrutiny of the formative political-cultural struggles that took place prior to the mid- and late nineteenth century (the starting point for Yashar's in-depth analysis of Guatemalan and Costa Rican development). For as I demonstrate below, the political-cultural legacies of the colonial period and the political-cultural adaptations of the postcolonial years decisively shaped the developmental trajectories of Nicaragua and Costa Rica throughout the remainder of the century.

This article therefore offers an explanation for identity formation and power contestation under the colonial regime; it also spells out the impact of collective memory and rhetorical practices on national identities and the formation of political and economic structures in the postcolonial era (1821–50). Standard approaches to Latin American political history tend to overlook these dynamics, however.[43] This is a

[41] Paige, *Coffee and Power: Revolution and the Rise of Democracy in Central America* (Cambridge: Harvard University Press, 1997), 47.

[42] Yashar, *Demanding Democracy: Reform and Reaction in Costa Rica and Guatemala, 1870s–1950s* (Stanford, Calif.: Stanford University Press, 1997).

[43] Ralph Lee Woodward, for example, astutely concentrates on familial and localist struggles but does not connect them to actors' rhetorical strategies. See Woodward, "The Aftermath of Independence, 1821–1870," in Leslie Bethell, ed., *Central America since Independence* (New York: Cambridge University Press, 1991). Similarly, Victor Bulmer-Thomas treats the emergence of national identity in Latin America as dependent on the success or failure of economic development. Bulmer-Thomas is certainly correct that the two are related. But as I argue below, the causality runs in the other direction: national identity shapes developmental outcomes. See his excellent work, Bulmer-Thomas, *Economic History of Latin America since Independence* (New York: Cambridge University Press, 1994), 19–45.

serious blind spot. Postcolonial Costa Rican political leaders, for example, increasingly relied on the notion that their colonial experience had bequeathed to them a tradition of civility that defined their mandate for the future—to protect their unique way of life from *external* disruption. Nicaraguans, by contrast, came to believe that they were threatened from *within* by evil and anarchy. Both beliefs derived from critical rhetorical battles and settlements that shaped group memories and local power relations under the colonial regime (Sections III and IV below). Yet these beliefs did not develop ipso facto into coherent national identities. For this to happen, the legacies of the past and the political agency of those living in the moment had to come together. Legacies and agency were joined in the rhetorical strategies of competitive postcolonial elites (Section V). These strategies—and the rhetorical frames they entrenched—can be summarized as the Alchemists' Trick and the Fatalists' License mentioned earlier. But to understand these frames, we must first grapple with the rhetorical struggles, the power relations, and even the historical accidents that went into the creation of collective memory during the long colonial epoch. It is to this task that I now turn.

III. The Rhetoric and Politics of Conquest

If ever there was a society of political entrepreneurs, it was colonial Spanish America. Denied the right to elect representatives to the *córtes* in Spain, colonials were granted the privilege of corresponding directly with the royal sovereign. In fact, the crown charged its overseas subjects with the task of keeping the metropole "properly informed."[44] King Ferdinand decided early on that he would not tolerate a governor's interference with the mail because "all who want to write to us should be free to do so, and any information received we will inquire about and we shall make our decisions on the basis of the whole truth; and truth once known we shall decide."[45]

Not surprisingly, letters to the king proliferated, and from the start conquerors exhibited a "story-making predilection."[46] The story at the core of these letters, however, was nothing less than a rhetorical strat-

[44] J. Crow, *The Epic of Latin America* (Berkeley: University of California Press, 1992), 167.

[45] Ursula Lamb, "Cristobal de Tapia v. Nicolás de Ovando, Residencia Fragment of 1509," *Hispanic American Historical Review* 33 (August 1953).

[46] Inga Clendinnen, "'Fierce and Unnatural Cruelty': Cortés and the Conquest of Mexico," in Stephen Greenblatt, ed., *New World Encounters* (Berkeley: University of California Press, 1993), 15.

[illegible] Even as the details of the letters v [illegible], and protagonists, the central strategy wa[s] [illegible]. Adhering to the rhetorical frame of the conquest—a Manich[æan] frame that pitted virtue against vice—political entrepreneurs sou[ght] not only rewards for themselves and their allies but also punishment [for] their rivals. Conquerors, officials, and friars thus strove to convince t[he] royal arbiter that while they were "dutiful" and "truthful," their riva[ls] were "avaricious," "cruel," "disagreeable," "evil," "artful," and above all, "disloyal" to God and crown.[47]

Political entrepreneurship, rhetorical power, and identity formation also came together in the metropole. As a critical entrepreneur, for example, the friar Las Casas managed to persuade the sovereign that the *encomienda*—the system whereby the crown "commended" whole groups of Indians to "worthy" conquerors, who in exchange were awarded the tribute and labor of their charges—was so abused in the Indies that it eroded the legitimacy of imperial conquest and compromised the crown's moral authority. This was an argument that the crown, jealous of its exemplary identity, could not ignore. Through the New Laws of 1542, the crown launched a protracted but ultimately mortal attack on the *encomienda* system.[48]

Although conquerors everywhere were scandalized by the New Laws, their adaptive responses were varied.[49] In Nicaragua (as in Peru) they staged a rebellion. In most other colonies, however, rebellion was averted. The variation had much to do with the way rhetorical conflicts were settled. In Guatemala, for example, conquerors hammered home the point that the New Laws were the handiwork of clever friars and avaricious officials who used their formidable powers of persuasion to "dupe" the sovereign and his council.[50] These officials and friars, not the conquerors, were the real threat to the colonial universe and within that universe to the "Republic of Christians."[51] The Guatemalans' efforts paid off. A stack of royal decrees indicates that the crown responded

[47] Irwin Blacker and Harry Rosen, *The Golden Conquistadores* (Indianapolis: Bobbs-Merrill Company, 1960), 61–62.

[48] Ibid., 161.

[49] See Claudio Veliz, *The Centralist Tradition of Latin America* (Princeton: Princeton University Press, 1980), 29.

[50] James Lockhart and Enrique Otte, eds., *Letters and People of the Spanish Indies, Sixteenth Century* (NewYork: Cambridge University Press, 1989), 73–77.

[51] For cabildo letters, see Javier Ortiz de la Tabla et al., eds., *Cartas de Cabildos Hispanoamericanos, Audiencia de Guatemala* (Letters from Hispanic-American Cabildos, Audiencia of Guatemala) (Seville: Publicaciones de la Diputación Provincial de Sevilla; Publicaciones de la Escuela de Estudios Hispano-Americanos, 1984).

forcefully when informed of anomalous conduct on the eaucrats and religious orders.[52]

uatemalan conquerors, in fact, managed through these rhetores to extract exemptions from the crown, diluting and postthe local effects of the New Laws. In Nicaragua the outcome dically different. As per the New Laws, the high court in emala ruled that the governor of Nicaragua must give up his *ennda*. He soon returned to Spain to press his case (1548). But the nor's troubles, significantly, were the result of a rhetorical campaign that the bishop of Nicaragua had been waging for some time against the official and his clan. In his first report to the king, in 1544, the bishop accused the governor and his clan of comporting themselves as if they were "lords of everything." The bishop's subsequent letters also expressed concern over the maltreatment of Indians. On the latter point the bishop reminded Charles V that one day he—the king—"would have to account to God." The bishop then joined forces with Las Casas: in a report to the Council of Indies, they argued that the "tyranny" and "violence" of Nicaragua's ruling clan imbued Indians with hatred and disdain for Christianity.[53]

These authoritative condemnations triggered a series of rhetorical countermoves. First, Nicaragua's *encomienda* holders argued (as all conquerors tended to do) that the laws violated their contract with the crown. Second, they insisted that the laws had been "inspired" by "wicked" friars. Third, they supported their particular grievances with reference to a general colonial view that held the king above reproach while imputing evil intentions to his officers (and even friars and ecclesiastical hierarchs). With the justification for rebellion finally in place, the governor's sons could take action while their father the governor was still in Spain: they stabbed the bishop to death in 1549.[54]

[52] For the original texts, consult Federico Arguello Solórzano and Carlos Molina Arguello, eds., *Monumenta Centroamericae Histórica: Colección de Documentos y Materiales Para el Estudio de la Historia y de la Vida de los Pueblos de la America Central* (Monumenta Centroamericae: Collection of materials and documents for the study of the history of the peoples of Central America) (Managua: Instituto Centroamericano de Historia, Universidad Centroamericana, 1965).

[53] Frances Kinloch Tijerino, "La cruz frente a la espada," in *Revista de Historia*, no. 5–6 (1995), 11–14.

[54] The absence of Nicaragua's governor from the province was a crucial historical accident, in the sense that with the governor away in Spain, his sons and their allies had no high representative of royal authority to turn to for on-site mediation. The only other high representative of the crown was the bishop—the very embodiment of evil in the rebels' eyes. Had the governor remained in Nicaragua, one of two alternative scenarios would likely have materialized. Either the governor himself would have led his clan in rebellion against the laws, in which case the outcome would have been even more similar to that of Peru. Or the governor and his clan would have refrained from rebellion and opted instead

From the perspective of the crown the rebels had confirmed La
Casas' most disturbing argument against the *encomienda:* the coloni
regime, controlled by tyrannical governors and abusive *encomiend*
holders, was sliding into moral and political illegitimacy. The Nica
raguans' *open* repudiation of the New Laws, in fact, rendered the enter
prise of conquest yet more vulnerable to the political, theological, an
juridical attacks that were by then current in Europe.

In Spain itself the Indian question became a matter of bitter contro
versy, accentuating the split between those who believed hierarchy wa
the natural state of human society and those who saw equality as its in-
herent condition.[55] But the emergence of a new vision in the field o
imaginable possibilities was at no point more palpable than during th
famous polemic of 1550–51 between the humanist Sepúlveda and th
friar Las Casas. On one side, the humanist argued that Indians an
Spaniards were as different as "men and monkeys." Indians, in thi
view, were simply beyond redemption. On the other side was La
Casas, who argued that Indians, no less than Spaniards, were childre
of God, thus rejecting the notion of an insuperable moral gap betwee
the races.[56] According to this alternative view, while the world was di-
vided between morally "backward" and more "advanced" peoples, re-
demption was a *universal possibility.*

The crown ultimately assumed a synthetic position in this contro-
versy by commissioning what was in effect the Americas' first officia
story. The state's senior cosmographer and chronicler was charged wit
the task of crafting a metanarrative of conquest, for the purpose of con-
firming "the justice and right of the Spanish crown's claim to domin-
ion." Completing this task entailed a series of key rhetorical moves. Th
first move was to admit "the abuses" of the conquerors while emphasiz-
ing the crown's efforts "to protect its native subjects." The second mov
was to praise Las Casas' labor on behalf of Indians, so as to valorize th
support given the good friar by the monarchy. The third move was t
depict those same Indians as "great liars, drunkards, thieves, and i
some cases, sodomites," so as to underscore the urgent need for th
"good government" that only Catholic Spain could provide.[57]

in favor of de facto dilution of the laws, making the outcome similar to that of Guatemala. In these al
ternative scenarios, the logic of Manichaeanism would still have entered into play. But the bisho
might not have been the target of the rebels' Manichaean outrage, and the "sacrilegious" dimension o
identity-formation might have been precluded.

[55] Tzvetan Todorov, *The Conquest of America* (New York: Harper Perennial, 1982), 149–53.

[56] D. A. Brading, *The First America* (New York: Cambridge University Press, 1991), 82–100.

[57] Ibid., 209–10.

IV. Memory and Power

The state's official story, finally published in 1601, was a rhetorical summation that in essence closed off the possibility for redemptive rhetorical campaigns on the part of Nicaraguan colonials. Memory formation and power relations were profoundly affected by this turn and so was the construction of the province's self-image. To begin with, the crown punished disobedient Nicaragua by eliminating its governorship and placing it under the direct rule of the Guatemalan authorities for over twenty years. This was crucial because the province's political subjugation, in combination with the devastation of the Indian population at the hands of the province's ruling clan, rendered the province unappealing to potential settlers. Spaniards came and went, disillusioned by the grim prospects for personal enrichment and political power. Consequently, Nicaragua was quite slow to develop a stable creole elite.

Second, the absence of such a stable elite meant that for a long time Nicaragua had no significant political actors with any durable interest in displacing or modifying the memories of anarchy and tyranny associated with the conquest of the province. Third, by the time Nicaragua was released from the tutelage of the authorities in Guatemala and a local elite slowly began to form, the crown had published its official story, bringing to an end the state's formal process of incorporating and excluding memories from recorded history.

The upshot of all this was that the Nicaraguan colonials of the seventeenth century had little incentive to engage in rhetorical campaigns. Thus, no record was made of postrebellion Nicaragua, no account given of its growing immersion in the normalcy of a poor and marginal colony. All that was left for posterity were the older accounts of Nicaragua's founding moment; all that was left were the horrors perpetrated by the province's conquerors against local Indians, the conquerors' sacrilegious murder of the bishop, and the province's fall from royal grace. Subsequent *uses* of this record by national scholars and intellectuals deepened the impression that regardless of the angle from which one examined the province's recorded past, the forces at play were the same: ambition, despotism, and violence, all combining and recombining to render Nicaraguans as tyrants, anarchists, or both.[58]

[58] The first "official" history of Nicaragua, written by a conservative in 1882, depicted Nicaragua's early conquerors as cruel and ambitious and the bishop whom they stabbed to death as a pious man. In 1889 an anticlerical liberal historian painted the bishop as a "victim of his own arrogance." But another liberal historian soon restored the initial interpretation, emphasizing the vileness of Nicaragua's conquerors while vindicating the bishop as a staunch protector of Indians. Tijerino (fn. 53), 5.

For illustrations of how the colonial historical record, taken prima facie, shapes the historiography

This last interpretation not only prevailed but over time took on an aura of self-evident truth, in good measure because scholars and intellectuals failed to problematize its construction and deployed it for political-didactic purposes.

By contrast, the New Laws, so fundamental in shaping Nicaragua's ignoble image, had a fortuitous effect on Costa Rica's process of memory preservation and exclusion. Recall that the laws were issued in 1542, at which time Costa Rica's conquest was suspended between two seemingly discontinuous phases. The first phase peaked about 1539, when conquerors managed to penetrate Costa Rica's interior. That phase, however, soon came to a close because the savage persecution of Indians provoked an uprising, which in turn beat the conquerors into retreat. The second phase began only in 1561, almost two decades *after* the New Laws had been promulgated and upheld in the face of the Peruvian and Nicaraguan rebellions. So the conquerors arriving in Costa Rica would be forced to operate under changed conditions, which, though severely constraining at the time, allowed for better image making.

On one side the crown now saw clearly that conflict among conquerors and extreme violence against Indians were counterproductive. Accordingly, it instructed conquerors to "govern [themselves] in peace and tranquility" and to deal "peacefully" with the Indians, who were to be "attracted" to the faith.[59] On the other side were the conquerors of the second phase; acquainted as they were with the lesson learned from the cautionary tales of Nicaragua and Peru, they could expect to pay a heavy price for openly disregarding the New Laws.

Against this backdrop the province's founding governor, Juan Vázquez de Coronado, took special care when drafting the reports that would become the regnant theme of Costa Rica's colonial memory. Most notably, he followed very closely the new royal guidelines instructing conquerors to live in harmony with one another and to "attract" Indians to the faith. He also studiously avoided the vitriol that other conquerors had poured out in their reports to the king when relating the behavior of fellow Spaniards in the colonies. He dwelled on the "great love and benevolence" with which *he* treated Indian chiefs,

---

of the nineteenth and twentieth centuries, see Tomás Ayón, *Historia de Nicaragua desde los tiempos mas remotos hasta el año 1852* (History of Nicaragua from the most remote times until the year 1852) (Managua: Tipografía de el centro-Americano, 1889); José Dolores Gámez, *Historia de Nicaragua desde los tiempos prehistoricos hasta 1860* (History of Nicaragua from prehistoric times until 1860) (Managua: Tipografía del país, 1889); Alemán Bolaños (fn. 32); Coronel Urtecho (fn. 32). See also José Coronel Urtecho's introduction to Ernesto Cardenal, *El Estrecho Dudoso* (The doubtful strait) (Mexico and Buenos Aires: Ediciones Carlos Lohlé, 1972).

[59] Meléndez (fn. 30), 41–47, 49.

but he also downplayed intra-Spanish conflict.[60] And he enveloped his narratives in an almost technical aura, as when he very straightforwardly reported that his benign method of colonization proved "so efficacious that it was publicized throughout the land, and [that one Chief] . . . came to render obedience and to recognize his serfdom."[61]

These exemplary exercises in reportage, then, documented the ideal conquest —a conquest seemingly free of jealousies among conquerors and above all a conquest apparently accomplished without recourse to willful deceit or wanton violence against the natives. But there was more to rhetorical politics than the effectiveness of these particular reports. Several additional factors stand out. To begin with, this idealized rendition of the conquest went *unchallenged* because Costa Rica had no professional chroniclers, which meant that the rhetorical contentiousness that such professionals had introduced in prior conquests was now absent. Furthermore, there were no articulate defenders of Indians in the mold of the friar Las Casas or the bishop of Nicaragua. That is, there was no one to raise authoritative criticism when colonials deviated in practice from the letter of their self-serving reports[62] and/or disregarded the wishes of the crown.[63] The result was rhetorical hegemony at the critical point of Costa Rica's "founding" moment: the leading conqueror not only controlled the flow and substance of reports from the periphery to the metropole, but just as importantly his reports created an impression of *general* harmony.[64]

The impact these accounts had on the authorities in Spain, though important, was only secondary in significance to their impact on Costa Ricans themselves. Most notably, the unchallenged accounts provided the province's founding "nobles"[65] with an intangible but crucial resource: a good reputation. The nobles both exploited and cultivated this resource. They carefully observed—if only in appearance—the crown's increasingly sharp admonitions against violent conquest, most clearly

[60] José Francisco Trejos, ed., *Progenitores de los Costarricenses, Los Conquistadores* (The Costa Ricans' progenitors: The conquerors) (San José: Imprenta Lehmann, 1940), 34.

[61] For the original texts, see *Cartas de Relación Sobre la Conquista de Costa Rica* (Narrative reports on the conquest of Costa Rica) (San José: Academia de Geografía e Historia de Costa Rica, 1964).

[62] If anything, the first Franciscans to come to Costa Rica served as the "conquerors' right hand." Arnoldo Mora Rodríguez, *Historia del Pensamiento Costarricense* (History of Costa Rican thought) (San José: Editorial Universidad Estatal a Distancia, 1992), 64.

[63] There was, for example, no such controversy when the Costa Rican governor, a most obedient subject by his own account, actually ignored a royal ban on the allotment of Indians (a ban that forbade conquerors from distributing Indians among themselves)

[64] Francisco Montero, *Historia de Costa Rica* (San José: Tipografía Nacional, 1892), 40.

[65] The crown seldom granted titles of nobility in Spanish America, but this did not diminish the nobiliary pretensions of Costa Rica's conquerors.

embodied in the Royal Ordinances of Pacification of 1573.[66] They extracted favors from the metropole on the grounds that their peaceful conduct left them exposed to the Indians' savagery![67] They staked a credible *common* claim to a privileged ancestral identity, which they then used to present a united front when dealing with Spanish officials. They played on the crown's high valorization of dutiful obedience to establish a de facto right to vocality in the selection of their Spanish governors.[68] And they asserted on the basis of (seemingly) meritorious conduct their own standing vis-à-vis governors, whom they eventually reduced to veritable subalterns.[69]

By the 1810s—a mere decade away from independence—Costa Rica's governors routinely pressed the nobles' agenda and accorded them the respect they craved; and the nobles, for their part, celebrated their governors as "protective fathers" whose "progressive" administrations promoted the province's tranquillity." Meanwhile, in Nicaragua (as in the rest of Central America) acrimonious relations between Spaniards and creoles destabilized the colonial order. The contrast merely emphasized Costa Rica's exceptional image—to the point that in a late-nineteenth-century retrospective it was said that while Costa

[66] Royal ordinances on "pacification," 1573, in Lewis Hanke, ed., *History of Latin American Civilization, Sources and Interpretations*, vol. 1, *The Colonial Experience* (Boston: Little, Brown and Company, 1973).

[67] Montero (fn. 64), 44.

[68] Ricardo Fernández Guardia, *Crónicas Coloniales* (Colonial chronicles) (San José: Imprenta Trejos Hnos., 1921), 517.

[69] For power relations in colonial Costa Rica, see Cruz (fn. 37), 85. Reputation building, of course, took place within particular institutional contexts. Costa Ricans, for example, had no institution of higher learning. (Any Costa Rican seeking enlightened cultivation had to study either in Guatemala City or in Nicaragua, at the seminary in León). This deficiency, though lamented by the nobles themselves, meant that no institutional framework was available for the elaboration and recording of narrative challenges to the nobles' exceptionalist account of conquest and colonization. In Nicaragua, meanwhile, the seminary helped entrench the "fall from grace" view of history. The seminary, for example, opened up to the novel philosophies and political theories coming out of Europe and the United States but then molded them to the Christian worldview. Leon's "sacred orators," for example, strove in the late eighteenth century and the early decades of the nineteenth to reconcile the progressive spirit of the Enlightenment with the immutable truths of the Christian worldview. And even as independent intellectual clubs began to form in the late nineteenth century, the church's influence remained sufficiently strong to extirpate revisionist interpretations of the bishop's murder. The short-lived counternarrative of the murdered bishop as victim of "his own arrogance" and the restoration narrative in which the bishop once again figured as a victim of "vile conquerors" were both penned by liberal notables associated with the city of León. This brief interpretive conflict and its swift resolution reflected the long-standing institutional entwinement of León's educational elites and the church qua guardian of hierarchical orthodoxy. For early examples of this entwinement, see "Discurso," pronounced at the León seminary, May 16, 1807. See also "Sermón," delivered by Tomás Ruiz, professor of philosophy and vice-rector of the seminary, November 27, 1804; and "Sermón en los Funerales del Padre Don Rafael Ayesta," delivered by Francisco Ayerdi, both reproduced in *Boletín Nicaraguense de Bibliografía y Documentación*, no. 55 (November 1987/September 1988).

Rica had been the crown's most neglected territory, in no other had the metropole been "better loved."[70]

This seemingly fluid and happy memory, as we have seen, was grounded in a particular process of memory construction and reputation building that produced a power structure in which royal officials became the local elite's junior allies. In the next section we will see that, paradoxically, it was precisely this happy memory that introduced *competing* notions of realism into Costa Ricans' postcolonial field of imaginable possibilities.

V. Alchemists and Fatalists at Work

The keenly competitive political leaders who emerged on the scene in postcolonial Costa Rica (1821) were quintessential political entrepreneurs. But they were soon in a bind precisely because—unlike the Nicaraguans—they perceived more than one possible route to power. One available option was simply to follow the Central American example; that is, Costa Rican leaders could compete with each other by riding high on the tide of localism. This was the politically easier but bloodier route because it called for each rival leader to try to impose the predominance of his respective locality at the expense of the others. The other option was less costly in terms of lives but was more complicated in a context of heightened localism: rival leaders could compete for political dominance by exploiting Costa Ricans' colonial self-image as a poor but peaceful people. Costa Rican political entrepreneurs experimented briefly with both options, one moment jockeying for localist preeminence, the next proposing that all localities band together behind one locality in order to keep at bay the "fratricidal" wars convulsing the rest of the isthmus.

Neither option prevailed; just as one gained relative salience the other recovered strength. In turn, this unresolved competition engendered ambivalence among political entrepreneurs—an ambivalence they transcended through intense contestation, which then produced yet a third alternative. Specifically, entrepreneurs set out to conciliate two opposing objectives: assuring collective insulation from external disruption *and* vindicating internal localist claims. Eventually, they

[70] Francisco Montero Barrantes, *Elementos de Historia de Costa Rica* (San José. Tipografía Nacional, 1892), 153. And Francisco María Iglesias, "septiembre 15" (San José, September 15, 1888), in *Revista de Costa Rica en el Siglo XIX* 1 (San José: Letras Patrias no. 32); and Máximo Soto Hall, "Capítulos de un Libro Inédito," in *Revista de Costa Rica en el Siglo XIX* 1, 87.

came to see reconciling of these two objectives as the sine qua non of political success. Indeed, the imperative for conciliation proved binding because, by operating within the parameters of what they thought was realistic, political entrepreneurs further reconfigured their collective field of imaginable possibilities.

This process of reconfiguration began when competing leaders actually managed to forge a foundational compromise known as the Pact of Harmony (1821). One of the pact's crucial provisions was the institutionalization of an itinerant national government: high officials were obligated to reside in each of the country's four major cities for part of every year. The peculiar arrangement did something unprecedented in the rigidly hierarchical colonies: it divided up equally the prestige and power attached to the seat of government. The argument advanced in favor of this curious arrangement was derived, ironically, in part from Costa Ricans' understanding of their unique colonial experience, as, political entrepreneurs concluded that Costa Rica's long tradition of internal harmony could be preserved from external threat only if all domestic localist claims were satisfied in some equitable fashion.

The arrangement was also fragile, of course. In 1823 and 1835 there were armed encounters between rival localities, with each trying to establish its own predominance by destroying the arrangement itself. But these conflicts were also very quickly resolved. Hostilities in 1823, for example, lasted no more than a few hours, while similar localist conflicts went on for years in the rest of the region. More important still, the pact that encoded the arrangement proved quite resilient. In fact, over time, the Pact of Harmony became the emblem of Costa Rica's exceptionally peaceful transition to the national era, even taking on a numinous quality not unlike that of the U.S. Constitution.

To understand the pact's resilience—and its beneficial effects—we must now turn our attention to the rhetorical practices associated with the declarative process of identity formation. These rhetorical practices hinged on a series of "realistic" assertions about Costa Ricans' place in a dangerous world made by political entrepreneurs trying to capture allegiances. These assertions emphasized their people's unique capabilities and weaknesses. In 1822 the country's first governing junta declared peace to be "virtually innate" to Costa Ricans because their colonial ancestors had adhered to a "sacred religion" that consisted of a "characteristic unity of opinion and calm."[71] The first chief of state (1824–33)

[71] "Proclama de la Primera Junta Superior Gubernativa al Terminar su Período," Alajuela, November 9, 1822, in Carlos Meléndez Chaverri, ed., *Documentos Fundamentales del Siglo XIX* (Fundamental nineteenth-century documents) (San José: Editorial Costa Rica, 1978), 93–94.

helped bolster this assertion, as did the State Assembly.[72] The legislature declared that while Costa Ricans were "poor," riches were in store for them because they were "industrious," "calm," and "patriotic." And reflecting deep concern over Nicaragua's relentless upheaval, the assembly proclaimed in 1825 that if Costa Ricans guarded against the "corruption of their mores," they would enjoy "eternal and exemplary tranquillity."[73] For his part, the executive vowed to "draw" on the "virtues, morality, and good sense" of his compatriots. Only distance, he told them, would allow their small and weak country to maintain "public peace." Only peace, he added, could afford them "the time" to devote themselves to "business." And only in distant peace, he concluded, could they enjoy the "fruits of their labor and investments."

Thus, while in the rest of Central America clashes between these two branches of government over institutional predominance often led to war, in Costa Rica the branches outdid each other in their praise of the Costa Rican "people." But they did more. In 1825, they set out to provide the country's "virtuous people" with land via their first agrarian reform. This decision to begin distributing uncultivated tracts turned out to be the first step in what would become a concentrated effort to promote the production of coffee: from 1840 on, governments pursued a policy of free land distribution for that purpose.[74] But the reform was also meant to demonstrate the capacity of "humble" Costa Rica to defy expectations and single-handedly preserve its reputation and integrity. In 1827 the Costa Rican government declared that

> Costa Rica, which at every point has given unequivocal proof of its love for peace, order, and law . . . contemplates in horror the rapid degradation of the [Central American] Republic . . . the virtuous people of Costa Rica, notwithstanding their territory's small scale and the obscurity in which they have lived, and in spite of the erroneous and disdainful opinion about them held by others, have proved *capable* at the most critical points of surviving and retaining a *sense of self* without subsidy or assistance from their [Central American] siblings.[75]

This optimistic assessment proved correct, in no small measure because the strategy of matching factors of production—"virtuous"

[72] Astrid Fishel Volio, "La Educación en el Proceso de Formación y Consolidación del Estado Costarricense," in Gómez et al. (fn. 35), 129–52.

[73] "Manifiesto de la Asamblea del Estado," San José, April 19, 1825, in Meléndez Chaverri (fn. 71), 135–36.

[74] Torres Rivas (fn. 34), 17.

[75] Joaquín Bernardo Calvo, "Nota del Ministro General del Estado de Costa Rica al Ministro de Relaciones Interiores y Exteriores de la Federación," San José, October 8, 1827, in Meléndez Chaverri (fn. 71), 147–50, emphasis added.

human capital with uncultivated lands—paid off. By 1829 the executive was able to report that the country had managed to make steady progress in commerce, agriculture, and mining, while "enterprises, profits, and consumer goods" multiplied.[76] But at the same time, there were ominous signs no one had anticipated. Most notably, by 1831 the administration had accumulated a budget deficit that, in the eyes of the opposition (concentrated in the city of Cartago), imposed an unbearable "burden" on all the people.

Anywhere else in Central America this harsh indictment, in combination with intense localist pressures, would have been sufficient reason for immediate confrontation. But the Costa Rican opposition, while localist in its biases, pushed instead for the creation of a special commission to investigate the nation's fiscal problem. Once formed, the commission itself faced a menu of options. It could opt, for example, to press for a straightforward fiscal austerity plan or to recommend a deepening of the state's extractive capacity. Instead, the commission took the tack of arguing that the best way to increase the state's resources was to make citizens "rich," on the grounds that the state's fiscal soundness was inextricably tied to society's "general prosperity."[77]

The reasoning behind this position is instructive. Fiscal impositions were deemed problematic because they "exasperated" a people that had long sacrificed to keep its domestic "tranquillity." This was key, since tranquillity was considered the "sole support" of Costa Rica's "political viability." The commission elaborated on this last point by noting that "discontent and upheaval" in Costa Rica were an "uncharacteristic" state of affairs, whereas peace and industry were the country's very essence.[78]

This last assertion represented a critical shift in perceptions. Once convinced that their poverty under the colonial regime was a fact of life, Costa Ricans now took decisive steps to lift their country out of its "misery." In crafting its developmental plan, the commission produced a statement that was as much a chart for the future as it was a declaration of faith in the Costa Rican people. Costa Ricans, the commission noted, were eager to contribute to the welfare of the "republic." Land, however, was useless without credit, and the Costa Ricans' "good disposition" was wasted without employment opportunities. Nor could the

[76] Juan Mora, "Mensaje del Jefe de Estado a la Asamblea," San José, March 1, 1829, in Meléndez Chaverri (fn. 71), 150–57.

[77] "Dictámen de la Comisión para recabar arbitrios," Cartago, August 15, 1831, in Meléndez Chaverri (fn. 71).

[78] Ibid.

nation move forward if its leaders' hopeful utterances lacked credibility with the people. "Without wealth or credit, we cannot make credible promises of happiness and political viability," the commission stated, before going on to dispense one key piece of advice: grant titled plots of land to the poor, especially along the coasts. By the commission's logic, only "good titles" could establish a "spirit of good faith among citizens," and only in this way could Costa Ricans avoid "fights" over "what belongs to whom, and who owes what." No entity other than the state could provide this essential public good.[79]

The commission's plan, no doubt, was also a trenchant critique of the government in power. But the critical point here is the following: the more Costa Rican politicians expressed themselves in this way, even for self-seeking purposes, the more they unwittingly tied each other to the controlling notion that Costa Rica's status as a regional exemplar must be preserved at all costs. As one leader put it:

> Costa Rica has set a good example in every respect for the other states of the [Central American] republic. It is neither reasonable nor just that Costa Rica should now lose the beautiful prestige that it has earned through its constant vigilance in the protection of harmony, its exact adherence to the law, and its continuous repudiation of the partisan rows and personalist ideas that have been in abundant evidence in the other states.[80]

Further, Costa Rican politicians fell increasingly prey to the fear of contamination implicit in this 1830 statement. By 1835 dread of a Nicaraguan "infection" was very much in the open.[81] This is not surprising, given that the partisan rows and personalist ideas that Costa Ricans had repeatedly declared to be alien to their character were nevertheless resurgent in the country. In fact, the leaders of the various Costa Rican localities were on the verge of battle over the issue of national predominance. Nevertheless, in 1835, as in 1823, this conflict was very quickly resolved, thanks to three factors. One was the preponderance of force enjoyed by the city of San José. Another was the conciliatory pressure exerted by the hallowed Pact of Harmony on rival politicians (thus precluding triumphalist intransigence on the part of San José). And the third, related factor is that prior to the outbreak of hostilities, opposition leaders had outlined a developmental plan rooted in widely held expectations about what the future could and should

[79] Ibid., 481-85.
[80] José María Esquivel, "Sobre la Disolución de las Costumbres," San José, May 15, 1830, in Meléndez Chaverri (fn. 71), 474-75.
[81] "Manojito de Flores", *El Josefino*, San José, August 11, 1835, in Meléndez Chaverri (fn. 71), 485-88.

look like. In so doing, they had established a focal point for a postconflict working compromise, and generated powerful incentives for victors and vanquished alike to resume the business of growing the economy and protecting the nation's defining principles.

These incentives, in fact, further accentuated the pressure that rivals and allies alike put on incumbent executives to take quick and effective developmental action: hence the overthrow of an inadequate chief (1838) and the feverish pace and extensive scope of his successor's initiatives. In addition to building roads and ports to support large-scale coffee production and trade, the new chief (Braulio Carrillo) established civil and penal codes, a postal service, and a national defense apparatus—measures that culminated in the self-confident decision to separate Costa Rica from the Central American Federation.[82] Moreover, Carrillo fostered broad-based agrarian development while entrenching the elites' socioeconomic dominance. Starting with his administration, national governments systematically pursued a policy of free land distribution for the cultivation of coffee. (They also maintained a pyramid-like distributional structure. By 1850, with the coffee boom in full swing, "three quarters of all major coffee growers were descended from just two Spanish colonial families, one of them that of the conquistador Juan Vázquez de Coronado himself.")[83]

These accomplishments were so impressive that Carrillo was emboldened to expand the limits of executive authority. Pointing to his own record of effective leadership, he declared himself ruler for life. But as Juan Linz has observed, performance-based legitimacy, unlike procedural legitimacy, is not self-reproducing. That is, if the ruler fails to keep his programmatic promises, then he is simply not needed. And if he does keep his promises, he may not be needed anymore. Given additional time in office, Carrillo might well have attained new and impressive goals. But having accomplished so many already, he made himself dispensable and was forced out of power in 1842.

And yet, overcoming internal cleavages, political leaders ultimately vindicated Carrillo as a historical figure. This vindication process reached its crowning point in the 1970s, when Carrillo was officially named "benemérito de la patria"—worthy and deserving of the fatherland. But the vindication process had begun already in the second half of the nineteenth century. Depictions of Carrillo as a special kind of authoritarian became more frequent, as he was described as a firm and pa-

[82] Montero Barrantes (fn. 70).

[83] Paige (fn. 41), 15.

triotic ruler under whose wise leadership Costa Rica "grew accustomed" to "self-government" and "complete independence." And his accomplishments and reputation came increasingly to be tied to the larger theme of Costa Rica's collective virtuousness.[84]

Given the Costa Rican leaders' emerging understanding of their postcolonial history, the exaltation of Carrillo is not surprising. Specifically, leaders came to see this history as a series of character tests, each a challenge that had called on Costa Ricans both to reveal and to defend their true collective self. At every point since independence, in this account, Costa Ricans had to prove themselves anew as a "hard-working, pacific people, dedicated to agriculture and devoted to commerce." And at every point they confronted the fact of Nicaraguan neighbors, "possessed of many good and brilliant qualities" but "generally indolent, accustomed to internal strife, and inclined to fight with neighboring states." So that while in its territorial disputes with Nicaragua, Costa Rica had invariably remained "patient," "sacrificing," and "conciliatory,"[85] it was also the case that from a purely defensive perspective Carrillo's strong rule had been "necessary."[86]

By 1850 this view of the past, with its clear descriptive and prescriptive lines, had become an integral part of the dominant rhetorical frame. This frame shaped rivals' rhetorical strategies and their competing programmatic initiatives. This meant that rivals explained their designs and justified their actions by making vehement reference to the constitutive elements of identity, engendering in the process a collective field of imaginable possibilities. Put another way, rivals embedded themselves in a structure that helped define realism while imposing normative constraints on actors' self-seeking pursuits. From the mid-

[84] Emanuel Thompson, *Defensa de Carrillo: Un Dictador al Servicio de América* (In defense of Carrillo: A dictator at the service of the Americas) (San José: Imprenta Barrase, 1945).

[85] Felipe Molina, *Memoria sobre las Cuestiones de Límites que se versan entre la República de Costa Rica y el Estado de Nicaragua* (Record regarding border matters under discussion between the Republic of Costa Rica and the state of Nicaragua) (Madrid: Imprenta de la Viuda de Calero, 1850). Also refer to Pedro Pérez Zeledón, minister plenipontentiary to the United States, *Informe Sobre la Cuestión de Límites de Costa Rica y Nicaragua y Puntos Accesorios Sometidos al Arbitraje del Señor Presidente de los Estados Unidos* (Report on issues regarding the border and points of access between Costa Rica and Nicaragua, submitted to the president of the United States for arbitration) (Washington, D.C.: Gibson Bros., 1887).

[86] Molina (fn. 85), 25–26. For Costa Ricans' perception of the Nicaraguan threat, see *Documentos relativos a la guerra contra los filibusteros*, Document no. 215. For an account in which Carrillo's "prodigious" and "broad" contributions to "the nation's progress" ultimately redeem him as a historical figure, see the constitutional scholar Jorge Sáenz Carbonell, *El Despertar Constitucional de Costa Rica* (Costa Rica's constitutional awakening) (San José: Libro Libre, 1985), 362–70, 378. For a complete defense of Carrillo, see Thompson (fn. 84). For praise of Carrillo as the "true enlightened despot" who, playing a role similar to "Rousseau's supreme legislator," begins the construction of the sovereign state in Costa Rica, see Mora Rodríguez (fn. 62),118–19.

nineteenth century until the democratic revolution of 1948, political entrepreneurs occasionally imposed and deposed rulers and often manipulated electoral processes. But throughout all of this, they were also subjected to particular standards of performance that simultaneously entrenched Costa Rica's "exceptionalism," minimized political conflict, and deepened the developmental ethos of competing leaders. Moreover, throughout all of this, elites implicated the citizenry in the production and legitimation of power relations, promoting the emergence of a public sphere in which rulers and the people celebrated—shoulder to shoulder—the nation's civic traditions and ultimately affirmed the legitimacy of state authority.[87]

Conditions could not have been more different in Nicaragua, where immediately after independence localist and personalist leaders sparked a civil war (1823–24, 1827)[88] so intense and engulfing in scope that it was characterized as a horrifying display of fratricidal butchery by the Central American government.[89] This conflict is intriguing for its sheer intensity and displays of barbarism. But even more puzzling is the fact that although neighboring Costa Ricans had begun to show that divisive jealousies could be controlled (albeit imperfectly), Nicaraguans seemed bent on spoiling possible solutions to their destructive power struggles. Why was this the case?

The most important reason for this outcome is also the most counterintuitive: because Nicaraguans had no "happy" collective memory on which to draw in a context of rampant localism, they were free from the ambivalence that competing possibilities engendered among their Costa Rican counterparts. And so while Costa Rican political entrepreneurs felt compelled to apply their powers of persuasion to reconciling the available options—something they ultimately managed by emphasizing the common virtues and capabilities of their people—Nicaraguan entrepreneurs faced no such imperative. This is not to say that the construction of Nicaragua's dominant rhetorical frame was ei-

[87] See Margarita Silva, "Las Fiestas Civico-Electorales en San José Y El Reconocimiento de la Autoridad de Los Elegidos," in *Revista de Historia*, no. 27 (January–June 1993).

[88] For the localist and clannish rivalries underlying this conflict, see José Coronel Urtecho, *Reflexiones sobre la Historia de Nicaragua: de Gainza a Somoza*, vol. 1 (Reflections on the history of Nicaragua: From Gainza to Somoza) (León: Editorial Hospicio, 1962); José Dolores Gámez, *Historia de Nicaragua* (Managua: Fondo de Promoción Cultural, Banco de América, 1975); and José Coronel Urtecho, "Paradojas de las Intervenciones de Valle y Arce en Nicaragua," *Revista Conservadora del Pensamiento Centroamericano* 28 (May 1972).

[89] *Revista de la Academia de Geografía e Historia de Nicaragua* 6 (April 1944), 181–85; and "Discurso" pronounced at the Ateneo, September 15, 1881, in *Revista Conservadora de Pensamiento Centroamericano* 2 (July 1961).

ther entirely open or entirely foreordained. It was not entirely open because actors experienced rhetorical freedom—in a particularistic and immediate sense—in part because their only stable referent was a colonial experience that presumably offered nothing but strife and "treachery." And it was not entirely foreordained because, as in Costa Rica, political leaders were keenly alive to shifting opportunities and made critical choices accordingly.

As in Costa Rica, so too in Nicaragua, the interplay of agency and legacies was in evidence from the start. Take the case of electoral competition. In Nicaragua, as everywhere else, elections were meant to resolve conflicting claims to national dominance. But instead elections merely raised the already high levels of mistrust, a counterproductive effect that reflected rivals' Manichaean expectations. Thus, rivals expected electoral mechanisms, if uncorrupted, to reward the virtuous with victory even as they perceived one another as "wicked men" bent on corrupting electoral processes and controlling electoral outcomes. Indeed, in the face of their rivals' "perverse" motivations, leaders typically concluded that elimination of the foe was a prerequisite for elections. And when elections—secured by outside mediation—finally took place, even national-level winners tried to modify adverse results at the local levels.[90]

Moreover, rivals understood that persuading followers to go to war prior to elections or otherwise risk the turmoil associated with ex post interference with electoral outcomes required a rhetorical strategy that would drive home the point that these were unavoidable measures, given the harsh realities of their world. To this end, political entrepreneurs shifted the target of rhetorical campaigns away from a now irrelevant royal sovereign to *El Público* (the public). Here, the rhetoric political leaders bandied about is instructive. Rivals invariably characterized one another's utterances as "slanderous" and "malicious fictions." They invariably depicted each other as "monsters," "bloodthirsty," and "denaturalized." And just as invariably, they ascribed to their actions an evil intent: to "bewitch" and "daze" the "naive" and "innocent" public with their eloquence.[91]

[90] Elections were first held for deputies and for the top executive positions of chief and vice chief of state soon after Nicaragua was "pacified" in 1825 through the diplomatic intervention of Manuel José Arce (with the backing of his Salvadoran troops). Significantly, one of the terms of the peace agreement was the exile of the popular caudillo Cleto Ordoñez, whose *extraordinary eloquence* was considered a destabilizing force.

[91] "Un Nicaraguense," "Relación del Origen y Progreso de la Revolución del Estado de Nicaragua," May 10, 1827, *Revista de la Academia de Geografía e Historia de Nicaragua* 2 (September, 1937), 43; "Proclama de Juan Arguello sobre los sucesos de Guatemala en 1826," Libro de actas municipales de León, 1826, *Revista de la Academia de Geografía e Historia de Nicaragua* 1 (1936–37), 69–70.

In other words, they exploited the old colonial belief that the "wicked"—artful and conniving—"duped" the sovereign.[92]

The rhetoric of political entrepreneurs was also self-directed. Rival leaders strove to persuade themselves and each other that the Manichaean worldview they propounded was not only descriptive but also prescriptive. And indeed, by describing the world in such starkly negative terms, they put in place equally dark rationales. Cleansing the government of internal suspects, for example, was justified on the following grounds. First, a legitimate election is invalidated when, ex post, an elected official engages in "diabolical plots." Second, such plots reveal a "depraved spirit." And third, once revealed, the truth empowers the "valiant sons" of the *patria* to rescue her from a "perfidious" incumbent.[93]

In making this case, competing entrepreneurs generated a vicious cycle: attacks bred counterattacks ad infinitum. By the 1830s Nicaragua seemed to confirm the pessimistic expectation articulated by one Central American notable:

> If no punishment is set for those who slander in newspapers, those who have been slandered retaliate in kind; nations become uninhabitable. From one libel we go to another. Vengeance leads to the spilling of blood; and political societies succumb to chaos, death, and horror.[94]

The 1840s in Nicaragua were indeed characterized by episodes of partisan/localist violence and petty despotism. But these episodes were also punctuated by efforts at political learning. Some leaders argued that decades of "mutual disdain" and "killing" ought to have taught Nicaraguans two lessons. The first concerned the true nature of republican government, which, as the argument went, ought to be properly understood as citizens coming together in perfect unity to resist the "wicked."[95] The second lesson concerned freedom of expression. Here the message was that the individual's opinion must be subordinated to the pursuit of the collective welfare, a priority that required orators and pamphleteers to eschew exploitation of the "people's naive simplicity."[96] These two lessons combined to reinforce the view of the "popular sovereign" as "ignorant" and "pliable," while investing Manichaeanism with

[92] Manuel Antonio de la Cerda, "A Ciudadanos Secretarios de la Asamblea Constituyente," León, May 10, 1825, and May 19, 1825. *Revista de la Academia de Geografía e Historia de Nicaragua* 1 (1936–37), 255–56.

[93] *Revista de la Academia de Geografía e Historia de Nicaragua* 6 (1944), 155–56.

[94] "Discursos pronunciados en el Congreso Federal de Centroamérica el año de 1826, por José del Valle," *Anales de la Sociedad de Geografía e Historia de Guatemala* 2 (December 1925).

[95] For a typical statement, see *Registro Oficial* (March 1845).

[96] See, for example, José León Sandoval's broadsheet *Al Público* (Managua), September 22, 1846.

the ideological authority of republicanism. The effect was predictably polarizing: political entrepreneurs claimed for themselves the identity of ideological sophisticates bent on extricating the country from its "backwardness," and on this basis they then set out to remove the "little pueblo" from the rhetorical influence of "intrigue mongers."[97]

By the 1850s the country's political leaders were beyond reconciliation. In fact, dispossessed of any unifying perception of their common capabilities or a binding vision of national integrity, they were even ready to deliver themselves into the hands of the adventurer William Walker. All they had was a Manichaean worldview that reaffirmed the spreading perception of Nicaragua as an ill-fated nation. And so, as one ardent liberal put it, the hope was that the foreigner might act as "the tutelary Angel of Peace, and the North Star of the aspirations of an afflicted people."[98]

High instability and profound despondence overwhelmed early attempts by state elites to engage in concerted developmental efforts. In a brief moment of relative quiescence, for example, the State Assembly recognized that land—an abundant resource—could "take on value" only if placed within a proper incentive structure (1832). Accordingly, the assembly reaffirmed prior constitutional commitments to *ejidal* (communal) lands, and gave *ejidos* to towns, in hopes that "public happiness," "rural industry," and a class of "proprietors" would all thrive and bring about an improvement of popular "customs" and a stronger state.[99] But these donations notwithstanding, the development of agriculture in general and the cultivation of coffee in particular took their own spontaneous (and in the case of coffee, very slow) course until the late 1850s. This was no accident. Political elites were able to organize the state and bring political stability to the nation only after the establishment of the Conservative Republic, which for thirty years incrementally enhanced intraelite consensus and bipartisan cooperation.[100]

[97] For a representative statement, see the broadsheet published by Granada notables, *Al Público* (Granada), July 11, 1848.

[98] Francisco Vigil, *Manuscritos auténticos compilados, Padre Vigil* (first published in Granada, 1930; privately reprinted in Managua, 1967), 119–22.

[99] Julie A. Charlip, "'So That Land Takes on Value': Coffee and Land in Carazo, Nicaragua," in *Latin American Perspectives* 26 (January 1998).

[100] Put another way, the late development of the Nicaraguan coffee economy was rooted in political-institutional factors. One seemingly compelling alternative possibility—that late development of the coffee economy was due to structural factors like preexisting landholding patterns—is challenged by recent scholarly research. Large tracts of indigenous *ejido* (communal) holdings, for example, did not obstruct the early spread of coffee production in Nicaragua. To the contrary, as Charlip (fn. 99) has shown, ejidal lands allowed for the ubiquitous subsistence economy that sustained the coffee economy's *seasonal* labor force (pp. 2–5).

The puzzle that arises at this point, of course, is the republic itself—a puzzle that I have considered in detail elsewhere.[101] Suffice it here to say that both the republic's viability and eventual implosion were closely related to significant but ultimately fragile adaptations in the polity's rhetorical culture and attendant shifts in the field of imaginable possibilities. One adaptation in rhetorical convention stands out. Presidential incumbents were expected to remove their voices from partisan and localist exchanges and were therefore denied vocality in the most explosive rhetorical spaces of the public realm. Indeed, presidents not only practiced but even "declared" silence on such matters, thus explicitly disassociating their formal (constitutional) authority from their informal (rhetorical) power.[102] This disassociation diminished traditional elite fears that the executive might combine authority and power in order to monopolize the support of the "popular sovereign" and, by extension, arrogate excessive prerogatives unto himself.

This change in rhetorical convention and practice brought about a shift in the field of imaginable possibilities. Specifically, as fears of excessive concentration of executive power were assuaged, power sharing came to be perceived by antagonists as a realistic proposition. This crucial shift, in turn, had concrete political consequences. For if power could truly be shared, then bipartisan agreements and supralocal arrangements were no longer understood ipso facto as ploys of opponents. Nor was cooperation among rival leaders and localities necessarily seen as a betrayal of principle and interests.

The proximate causes of the republic's implosion reveal the underlying precariousness of these important changes. To begin with, failure by even one incumbent to adhere to the rule of declared silence could compromise the republic's entire institutional framework. In the early years of the republic this danger was minimized by careful selection of presidential candidates. But as the republic appeared to consolidate, dominant elites moved to enlarge the candidate pool. In so doing, they also enhanced the chances of electing a wayward president. And indeed, the republic imploded in the late 1890s. Then for the first time a notable associated with the city of León rose to the presidency of the republic, which for decades had been led by notables from the rival city of Granada and its satellite towns. Leon's notables declared the election a historical vindication for their city. But other localist forces, also present within the republic's institutions, balked at this interpretation.

[101] Cruz (fn. 1).

[102] See, for example, "Manifiesto de S.E. el Presidente D. Fernando Guzmán a los Pueblos de la República," (n.p.: Imprenta del Gobierno, 1867).

Governance quickly degenerated into acrimonious disputations in which the new president himself was a vociferous participant. Soon enough the republic was torn apart from within by mistrust, mutual recrimination, and subversive plots.[103]

The revival of the Manichaean frame that accompanied and followed the republic's crisis shows something else, that an expansion of the field of imaginable possibilities based on silence, however explicit this silence may be, is not the same as expansion based on fundamental and explicit discursive shifts. Because the republic left no clearly articulated alternative vision of the possible, for example, new generations of political elites and historians of various partisan affiliations and ideological stripes concluded that the republic itself was an "anomaly" in Nicaraguan history. Indeed, the republic became the antithesis of historical "normality," with the normal state of affairs defined as a perpetual oscillation between "anarchy" and "tyranny." Moreover, within this calamitous field of imaginable possibilities, the best hope lay with a strong, enlightened leader who might emerge to impose on the intractable nation a common quest for "order and progress."[104] Dictators did emerge, of course. But they also generated opposition—opposition that often returned to the conquest for its terms of reference, as when in the 1970s the anti-Somoza poet Ernesto Cardenal wrote that Nicaragua's founding (and evil) conqueror had been its first "dictator" and the first to "promote progress" (for himself).[105]

The case of Costa Rica stands in sharp contrast. Costa Rican self-invention proceeded in stages, with each stage borrowing from the prior one. In the 1880s, for example, political elites successfully used postcolonial exceptionalist claims to build the Liberal Republic. As a result, from the early decades of the twentieth century to the turbulent 1940s the country's "republican tradition" and "exceptional character" served as focal points for reformist debates involving a wide range of issues—from democratization of the electoral system to stabilization of agrarian relations. Even the crisis of the First Republic—culminating in the Civil War of 1948— hinged on political actors' contending interpretations of how best to preserve the glories of the past. Moreover,

[103] See also Arturo J. Cruz, Jr., "Overcoming Mistrust: The Quest for Order in Nicaragua's Conservative Republic, 1858–1893" (Ph.D. diss., University of Oxford, 1997).

[104] This was the hope that led liberal intellectuals at the close of the nineteenth century to rally around José Santos Zelaya. And this too was the hope that in the 1930s led modernizing intellectuals, even some who later became renowned critics of the Somoza regime, like Pablo Antonio Cuadra, initially to support Anastasio Somoza García.

[105] Ernesto Cardenal, *El Estrecho Dudoso* (The doubtful strait) (Mexico City and Buenos Aires: Ediciones Carlos Lohlé, 1972), 75.

the rhetorical settlement imposed by the 1948 victors tied the legitimacy of the Second Republic to the "redemption" and "improvement" of the First Republic's legacies. Indeed, the victorious leader José Figueres explicitly labeled the turbulent period of 1942 to 1949 as an "aberration" in Costa Rican history—a rhetorical move with several important implications. First, as Figueres himself contended, even his own armed rebellion and extraconstitutional junta could be deemed justifiable only in a context of "abnormality." Second, as Figueres further argued, now that Costa Rica had returned to the normalcy of its "inherent" civility and its "natural" democratic course, any new attempts at rebellion and extraconstitutional rule would be deemed unacceptable affronts to the nation's identity.[106] This last point was particularly important given the controversial nationalization policies of the new regime (and the new leadership's need to justify the smashing of reform counterthrusts). Third, since the 1948 revolution legitimated itself by claiming continuity with the country's "best" legacies, political rivals were gradually forced to craft new rhetorical strategies that (re)integrated pre-1948 elements of collective memory—elements that were no longer contestable and thus had a stabilizing effect on post-1948 elite competition. (During the increasingly trying decades of the 1970s and 1980s, for example, politicians of various partisan affiliations appealed to the citizenry for political support and peaceful unity by [1] invoking the exemplary behavior of their "benign" conquerors; [2] rehearsing the nation's "exceptional" transition from the colonial to the postcolonial era; and [3] exalting the early postcolonial Pact of Harmony.)[107]

CONCLUSION

In this article I have argued that collective identity is a robust declarative statement that a group makes, under the pressure of collective memory and contextual forces, to itself and to others about its normative constitution and its practical competence when facing the world. A normative constitution mandates what a group cannot or will not do out of deference to its self-defining convictions and sense of viable justice. Practical competence, in turn, refers to the group's intersubjective assessment of

[106] "La orientación general de la Segunda República," José Figueres, opening speech, constitutional assembly, January 16, 1949; and José Figueres, *Cartas a un Ciudadano* (Letters to a citizen) (San José: Editorial Universidad Estatal a Distancia, 1980).

[107] For striking examples, see José Manuel Salazar Navarrete, *Oduber, Elección Presidencial de 1974* (Oduber, presidential election of 1974) (San José: EDARASA, 1978); and Luis Alberto Monge, "Proclama de neutralidad," Presentación hecha en el Teatro Nacional, November 17, 1983.

its common capabilities and weaknesses. A dominant rhetorical frame is the mechanism that connects the two, in war as in peace.

I have also argued that although political actors are endowed with significant powers of initiative, actors' own rhetorical struggles can destabilize their purposive aims and alter their legitimate claims to power. Such struggles can even bring about shifts in a particular field of imaginable possibilities. These shifts either restrict or enhance actors' perceptions of their "realistic" options as they go about the business of connecting their self-seeking pursuits to their group's beliefs and expectations. In postcolonial Costa Rica, for example, competing elites deployed the memory of a "happy" colonial experience in order to garner support for their own particularist agendas. In so doing, they diminished their own chances at particularist dominance and improved the prospects for national unity and a common developmental agenda. Conversely, Nicaraguans sought unity by deploying a Manichaean colonial narrative that opened up ample space for self-seeking pursuits but eroded the foundations for consensus building and sustainable developmentalism. These divergent outcomes call into question, among other things, Anderson's view of national-identity construction as a process of "fraternal" imagining that obtained uniformly across postcolonial Spanish America.[108]

Finally, I have argued that to understand these contrasting outcomes, we must identify the rhetorical frames and the fields of imaginable possibilities that shape political identities and structure political competition. Only in this way can we explain why political entrepreneurs sometimes cultivate preexisting suspicions and conflicts and at other times act as innovative conciliators. Which path political entrepreneurs follow is of momentous importance. For postcolonial Costa Rica it meant relative political civility and early economic progress. For postcolonial Nicaragua it meant political-economic ruin. In the latter, even now, rampant mistrust and leaders' preemptive attacks on one another undermine the country's incipient democracy—one moment paralyzing major governmental institutions, the next provoking the feverish rounds of public graft that deepen popular disenchantment. Yet in both cases the lesson for political analysts and practitioners is the same: *how* we remember shapes *what* we can imagine as possible.

This means that there are no quick fixes for the noxious sentiments and the barbaric strife that often accompany identity-driven clashes. Improving the chances for pluralist peace in conflict-ridden societies

[108] Anderson (fn. 26).

requires more than changing the leadership, reforming formal institutions, or developing schemes for coexistence, important as all of these can be. It requires making critical entrepreneurs of us all: discerning citizens capable of distinguishing between history and fate. This must be the work of the village school and the think tank, the historiographer and the visionary. In some cases patient and careful foreign assistance or even enforcement may be useful. But the objective is always the same: deny exploitative entrepreneurs the ring of truth.

THE CAUSES OF WELFARE STATE EXPANSION

Deindustrialization or Globalization?

By TORBEN IVERSEN and THOMAS R. CUSACK*

INTRODUCTION

IT is commonplace to argue that the increasing openness of national economies has meant growing economic insecurity. This insecurity once supposedly fueled demands for larger welfare spending as a form of insurance.[1] The rising tide of globalization, however, is now widely seen as a hinderance to a government's ability to meet these demands and even as a cause of government cutbacks.[2] An alternative view combines this "second image reversed" with a concern for the political power of labor and the left.[3] This revisionist perspective suggests that the challenges promoted by globalization when met by strong left-labor power within the domestic political system combine to produce a compensation strategy that entails a large and vibrant welfare state. This paper challenges both these views. Our argument, in short, is that most of the risks being generated in modern industrialized societies are the product of technologically induced structural transformations inside national labor markets. Increasing productivity, changing consumption patterns, and saturated demand for products from the traditional sectors of the economy are the main forces of change. It is these structural sources of risk that fuel demands for state compensation and risk sharing.

* An earlier version of this paper was presented at the 94th American Political Science Association Meeting in Boston, September 3–6, 1998. For many helpful comments and suggestions on this and a related paper we wish to thank Keith Banting, John Freeman, Geoffrey Garrett, Peter Hall, Bob Hancké, Peter Lange, Paul Pierson, Jonas Pontusson, Dani Rodrik, Michael Shalev, David Soskice, John Stephens, and three anonymous readers. We also thank Geoffrey Garrett for providing us with some of the data used in our analyses. Both authors express their appreciation to the Science Center Berlin and the Center for European Studies at Harvard University for their support.

[1] See David Cameron, "The Expansion of the Public Economy: A Comparative Analysis," *American Political Science Review* 72, no. 4 (1978); and Peter Katzenstein, *Small States in World Markets* (Ithaca, N.Y.: Cornell University Press, 1985).

[2] Dani Rodrik, "Why Do More Open Economies Have Larger Governments?" *Journal of Political Economy* 106 (October 1998).

[3] See Geoffrey Garrett, "Capital Mobility, Trade and the Domestic Politics of Economic Policy," *International Organization* 49, no. 4 (1995); and idem, *Partisan Politics in the Global Economy* (Cambridge: Cambridge University Press, 1998).

The transformation of labor markets in recent decades is revealed in a dramatic shift in the employment structure. The two traditional and, until recently, leading sectors of employment, that is, agriculture and industry, have everywhere contracted. In the early 1960s an average of 60 percent of total employment was in agriculture and manufacturing. In the next three decades this figure dropped by nearly half (see column A, Table 1). In the United States, for example, 5 percent of the working-age population lost employment in these sectors over the last three decades, whereas in countries such as France, Germany, Sweden, and Denmark, the comparable figure is 15 percent or more. In addition to cross-national variance, the speed of the process has also varied a great deal over time. Sometimes layoffs have occurred in a slow, steady trickle; at other times they have been quick and massive, resulting in headline-grabbing factory closings.

Individuals face significant risks as a result of these shifts. Those thrown out of a job or threatened by the loss of employment may find that the skills they have acquired are not easily transferable to other parts of the economy where employment may be expanding, namely, the service sector. Even where employment is available, a job outside one of the traditional sectors often entails a significant loss in income as well as the deprivation, at least in part, of pension rights, medical insurance, and other work-related benefits. For many, indeed, loss of employment in the traditional sectors entails complete removal from the active labor force.[4] As one scholar notes, a significant part of this change in the occupational structure is due to the entry of young people into service employment and the early retirement of older workers from the traditional sectors.[5] This change is confirmed by the dramatic reduction in employment activity on the part of older workers who have in one way or another been pushed into early retirement during the last few decades.[6]

Broadly speaking, governments have responded to the transformation of the employment structure in three distinct ways. First, some

[4] There has been, however, significant cross-country variation in exposure to these risks. Those countries that were relatively early "deruralizers" were confronted with less problematic conditions, which helped them cope with structural change. Among these conditions were small entering-age cohorts, relatively low female participation in labor markets, and relatively buoyant labor markets. All of these conditions eased the problems of structural change. For a discussion of the timing of "deruralization" and its consequences, see Gøsta Esping-Andersen, *Social Foundations of Postindustrial Economies* (Oxford: Oxford University Press, 1999), 24–27.

[5] Hans-Peter Blossfeld, "Is the German Dual System a Model for a Modern Vocational Training System?" *International Journal of Comparative Sociology* 33, nos. 3–4 (1993).

[6] Martin Kohli, Martin Rein, Anne-Marie Guillemard, and Herman van Gunsteren, eds., *Time for Retirement: Comparative Studies of Early Exit from the Labor Force* (Cambridge: Cambridge University Press, 1992).

TABLE 1
CHANGES IN THE EMPLOYMENT STRUCTURE
(CHANGES FROM 1962 THROUGH 1993 IN THE SHARES OF WORKING-AGE POPULATION).

	A *Loss in Manuf. & Agric.*	B *Gain in Priv. Services*	C *A Not Absorbed by B*	D *Change in Gov.t Civ. Serv.*	E *Change in Military*	F *Not Absorbed*
Australia	13.4	11.6	1.9	3.8	–0.2	–1.8
Austria	16.9	5.5	11.3	6.2	0.1	5.0
Belgium	15.2	7.7	7.5	4.2	–0.7	4.0
Canada	8.1	11.4	–3.3	6.2	–0.7	–8.8
Denmark	21.0	5.2	15.7	14.9	–0.7	1.6
Finland	27.6	4.8	22.7	8.5	–0.6	14.9
France	20.9	10.4	10.6	3.9	–1.3	8.0
Germany	15.4	7.0	8.4	4.2	–0.2	4.4
Italy	19.5	8.7	10.8	4.0	–0.6	7.4
Japan	13.6	16.0	–2.5	0.4	–0.1	–2.8
Netherlands	13.3	16.8	–3.5	1.3	–1.2	–3.5
Norway	15.8	9.3	6.5	13.9	–0.4	–7.0
Sweden	17.7	2.9	14.8	14.4	–0.4	0.8
Switzerland	17.1	20.7	–3.7	4.2	–0.1	–7.7
United Kingdom	15.6	10.4	5.4	1.1	–0.5	4.9
United States	5.3	13.3	–8.0	3.5	–1.5	–10.0
With Australia:						
Average	16.0	10.1	5.9	5.9	–0.6	0.6
St. Dev.	5.0	4.7	8.2	4.5	0.4	6.7
Without Australia:						
Average	16.2	10.0	6.2	6.1	–0.6	0.7
St. Dev.	5.1	4.8	8.5	4.6	0.4	7.0

have promoted employment in private services, often by deregulating product and labor markets and allowing greater wage dispersion. At the same time, the governments have used various forms of public insurance to compensate workers for the risks of having to find new jobs in services. The United States is the archetypal example of this strategy, but Canada, the U.K., and more recently the Netherlands share some of the same features. In the U.S., since the expansion of private sector service employment has exceeded the relative modest loss in the traditional sectors, employment rates have actually increased (as indicated by the minus signs in columns C and F of Table 1).

The second strategy is for the state to maintain extensive regulation of private services as well as a relatively compressed wage structure

while simultaneously expanding employment in public services. Countries that have engaged heavily in this sort of strategy, most notably in Scandinavia, also have generally managed to elevate the total labor-force participation rate. On the spending side, government consumption has risen substantially, often complemented by an expansion of the state's public insurance functions in order to compensate for the risks associated with often very large employment losses in the traditional sectors (see the numbers for Denmark, Norway, and Sweden).

Finally, there are those economies where heavy regulation of labor and product markets has hampered a major expansion of private sector service employment. At the same time the public sector has not been allowed to grow to any significant extent. In combination with the large losses in the traditional sectors, this has led to a tremendous reduction in employment possibilities for those formerly active. Examples of states that have taken this route include Germany and France, and much of the welfare effort in these countries has been geared toward ensuring a relatively orderly and secure exit from the labor market, mainly through early retirement. Limiting labor-force participation in this manner is expensive, and, depending on the severity of shifts in the occupational structure, state-sponsored labor-market exit is often supplemented by an increase in the state's insurance role (as in the other countries). This response therefore creates transfer- as opposed to consumption-heavy welfare states.

The three responses clearly resonate with Esping-Andersen's typology of welfare states (liberal, social democratic, and Christian democratic),[7] and we believe that labor-market institutions and partisan politics have played an important role in shaping these responses.[8] By focusing on these major shifts in the labor-market structure and the partisan responses to them, we thus point to a causal structure that can help make sense of one of the most influential contemporary typologies of welfare states. The main focus of this paper, however, is to convince the reader that growth in both transfers and government consumption—the two main components of welfare-state spending—can largely be explained as a function of the severity of internally driven employment losses in the traditional sectors, *not* by forces in the global economy. Precisely because the underlying causal logic defines the available courses of political action, and hence helps us to account for

[7] Gøsta Esping-Andersen, *The Three Worlds of Welfare Capitalism* (Princeton: Princeton University Press, 1990).

[8] See also Torben Iversen and Anne Wren, "Equality, Employment, and Budgetary Restraint: The Trilemma of the Service Economy," *World Politics* 50 (July 1998).

the observed variance in welfare-state forms, *getting the causal story right is important.* This is also important for the sake of understanding how the politics of the welfare state is likely to change in the future. Since the processes of globalization and deindustrialization have very different distributions in time and space, the pattern of welfare expansion (or contraction) should vary accordingly.

The remainder of this paper is organized into four sections. In the first we examine the arguments of two leading scholars, both of whom see the increasing integration of national markets into the international economy as the most powerful force affecting governments' commitments to welfare spending. We suggest that there is little empirical evidence to sustain such a position. We then introduce and outline the logic of our own argument, focusing on the consequences of the employment dislocations connected to these major shifts in the occupational labor market structure since the early 1960s. In the third section this argument is tested on data for fifteen OECD countries over a period of thirty-three years, followed by an analysis that defends our argument and findings against the charges that deindustrialization is a result of either government spending itself or globalization. We conclude with a discussion of why domestic, as opposed to international, forces have been ignored in recent research, and we point to several areas where future research could prove fruitful.

DISCOUNTING GLOBALIZATION

The argument that globalization leads to welfare state expansion rests on two causal mechanisms. First, trade and capital market integration is said to expose domestic economies to greater real economic volatility, which implies higher income and employment risks for workers. Second, greater labor-market risks are hypothesized to generate political demands for expansionary spending policies that will cushion and compensate people for such risks. Rodrik focuses on the effects of trade and explains the logic in the following manner:

> More open economies have greater exposure to the risks emanating from turbulence in world markets. We can view larger government spending in such economies as performing an insulation function, insofar as the government sector is the "safe" sector (in terms of employment and purchases from the rest of the economy) relative to other activities, and especially compared to tradables.[9]

[9] Rodrik (fn. 2), 13.

Garrett extends the trade openness argument to globalization more broadly, including growing capital market integration:

> Perhaps the most important immediate effect of globalization is to increase social dislocations and economic insecurity, as the distribution of incomes and jobs across firms and industries becomes increasingly unstable. The result is that increasing numbers of people have to spend evermore time and money trying to make their future more secure.[10]

Left governments are more responsive to popular demands for compensation than right governments, according to Garrett, and his emphasis on capital market openness is also novel. The trade openness thesis, however, has a long history in political science, going back to the seminal works of Cameron, Ruggie, and Katzenstein.[11] To our knowledge the trade argument has not been subject to any serious challenges, and it stands out as one of the important explanations for the rise of the welfare state since the Second World War. The role of capital market integration is more contentious because of the effects such integration may have on macroeconomic autonomy,[12] but it is a logical extension of the trade openness argument.

We find it surprising that the alleged linkage between international economic exposure and labor-market risks has not received more critical attention. Although it is undeniable that international market volatility increases labor-market risks, whether openness is related to risk depends on the extent to which international market volatility is greater than domestic market volatility. It is not sufficient, for example, to show that international price volatility, measured in terms of trade instability, is related to spending.[13] In addition, at least one of two conditions must obtain: (1) price volatility in international markets is greater than in domestic markets, and (2) trade concentrates more than diversifies risk.

There are no theoretical reasons to expect the first condition to hold, and trade theory does not make strong predictions about the second. Although trade concentrates risks to the extent that it leads to special-

[10] Garrett (fn. 3, 1998), 7.

[11] Cameron (fn. 1); John G. Ruggie, "International Regimes, Transactions and Change: Embedded Liberalism in the Postwar Economic Order," in Stephen D. Krasner, ed., *International Regimes* (Ithaca, N.Y.: Cornell Univerity Press, 1983); Katzenstein (fn. 1).

[12] See Fritz Scharpf, *Crisis and Choice in European Social Democracy* (Ithaca, N.Y.: Cornell University Press, 1991); and Paulette Kurzer, *Business and Banking* (Ithaca, N.Y.: Cornell University Press, 1993).

[13] See Dani Rodik, *Has Globalization Gone Too Far?* (Washington, D.C.: Institute for International Economics, 1997), chap. 3.
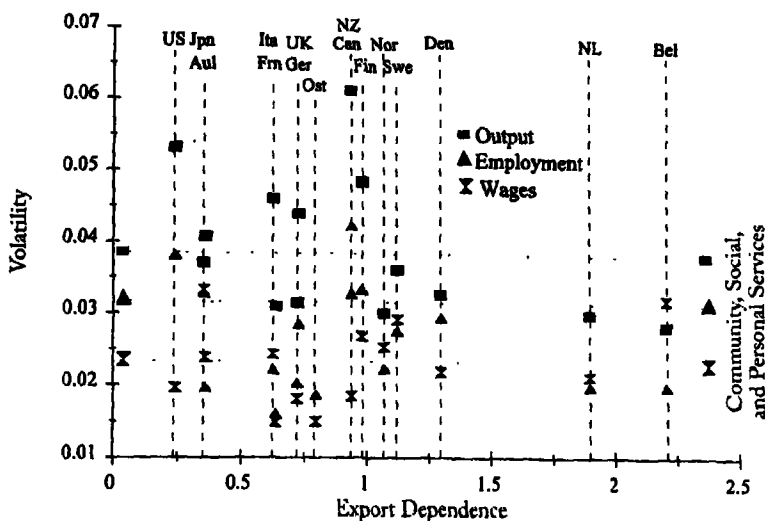
FIGURE 1
TRADE DEPENDENCE AND MANUFACTURING VOLATILITY[a]

SOURCES: OECD, *The OECD STAN Database* (1994).

[a] Export dependence is the total value of manufacturing exports divided by value added; volatility is the standard deviation in the rate of growth in manufacturing output, employment, and wages in the period 1970–93. Output data is not available for Austria; only employment data is available for New Zealand.

ization, it diversifies risks to the extent that it occurs across several national markets. Which effect dominates depends on the covariance of volatility across product and national markets. If specialization occurs within product categories that are exposed to similar cycles (complementarities) while trade occurs across national markets that are subject to different cycles, trade will actually lead to lower overall volatility. Since the bulk of trade within the OECD is intraindustry and occurs across numerous national markets, there is little a priori reason to expect that trade is associated with greater volatility. But only empirical evidence can resolve the issue.

For this purpose we have compared volatility in output, employment, and wages across the manufacturing sectors of sixteen OECD countries with very different exposures to trade (see Figure 1). Output and wages are measured in real terms, and volatility is defined as the standard deviation of annual growth rates between 1970 and 1993. This formula is

similar to the one used by Rodrik to measure volatility in terms of trade,[14] but here we are able to explore directly whether volatility in real variables is related to trade. As a baseline for the comparison, Figure 1 shows the average volatility of a completely nontraded (but private) service sector: community, social, and personal services (indicated by the three dotted horizontal lines).[15]

Contrary to the logic of the trade openness argument, there is no relationship between the export dependence of manufacturing (measured as the value of exports divided by manufacturing value added) and any of the volatility measures. The only variable weakly related to export dependence is output volatility, but the association is in the opposite direction of the one implied by the trade openness argument. Nor is there any evidence that the traded manufacturing sector is more volatile than the average for the nontraded service sector. Finally, it is noteworthy that there is no association between the level of volatility and Katzenstein's distinction between small corporatist welfare states and large liberal (or statist) ones.

If we changed the x-axis in Figure 1 to measure capital market openness instead of trade openness, the pattern would be no clearer. It does not appear to be the case that greater openness to the international financial system increases the volatility of the domestic real economy. Moreover, even if that proved to be the case, greater exposure to speculative capital flows may well be associated with a countervailing reduction in the capacity of governments to respond to pressures for compensation.[16]

But if these findings are correct, how is it possible that previous work has found such a clear link between globalization (especially trade openness) and spending? To answer this question we would like to draw attention to some important methodological issues in that work. Katzenstein is interested primarily in the effects of the global economic crisis in small countries during the interwar years, and he never presents any systematic evidence for the more general thesis that openness breeds compensation.[17] Focusing on small and open economies, it is

[14] Rodrik (fn. 2).

[15] The government sector *is* less volatile, but it does not make sense to include it in the comparison since this sector is supposed to be growing as a consequence of high volatility in the exposed sectors.

[16] This constraint could come on both the spending and the revenue side. Thus, there is broad consensus that capital market integration has raised the price of running deficits, while others (including Rod-rik) have argued that it has made taxation of capital more difficult. The latter argument, however, is disputed by Duane Swank, who finds no evidence that business taxation has become less important for revenue generation. See Swank, "Funding the Welfare State: Globalization and the Taxation of Business in Advanced Market Economies," *Comparative Political Studies* 46 (September 1998).

[17] Katzenstein (fn. 1).

difficult to assess whether expansion in the government's role in the economy is due to trade openness or some other features that these countries have in common. Cameron offers some cross-sectional evidence[18] but in the form of correlation coefficients or very simple regressions that fail to control for a number of factors (such as the size of the dependent population) that we now know are important. In fact, our data support Cameron's results in the sense that there is a cross-sectional association between openness and spending (r =0.6), but this relationship does not hold once proper controls are included in the statistical model (as we will show). In the case of Rodrik,[19] both cross-sectional and pooled time-series evidence is presented, but the analysis includes a large number of less developed and mostly nondemocratic countries to which our argument is not necessarily applicable.[20]

The results that are most relevant for our purposes are presented by Garrett.[21] Not only does Garrett focus on the same countries that we do, he also includes capital market integration in his analysis. Furthermore, Garrett's analysis picks up both cross-national and cross-time variance and allows for a large number of controls. It is therefore of considerable interest to replicate and further examine Garrett's results, as we have done in Table 2. The first two columns of the table replicate Garrett's results using change in government transfers and in civilian government consumption as the dependent variables.[22] First, note that the results for trade openness are weak and statistically insignificant. Somewhat surprisingly, Garrett's own results do not appear to support the trade openness argument. However, the coefficient for the interaction between what Garrett calls left labor power and capital market openness is positive and statistically significant, supporting his thesis that open capital markets lead to higher spending when the political left is strong and unions are encompassing. (Left labor power is a composite index of these variables.)

[18] Cameron (fn. 1).

[19] Rodrik (fns. 2 and 13).

[20] Indeed, trade openness may be more salient for less-developed countries because trade in many of these countries, unlike trade between OECD countries, has led to heavy dependence on a few primary commodities that are subject to high international price volatility.

[21] Garrett (fn. 3).

[22] Our thanks to Geoffrey Garrett for generously providing us with the data he used in his analyses, thereby allowing us to replicate his findings. See Table 4.4 in Garrett (fn. 3, 1998), 90. Note that Garrett uses levels of spending on the left-hand side, but this formulation gives an estimate for R-squared that is uninformative since the lagged dependent variable will pick up most of the cross-national variance. Using changes in spending on the left-hand side avoids this problem while leaving the estimated coefficients the same. (In mathematical terms, we are simply subtracting the lagged dependent level variable on both sides of the equal sign, which obviously leaves the coefficients for all other variables unchanged.)

TABLE 2
REPLICATING GARRETT'S REGRESSION RESULTS[a]

	Garrett's Results		*Our Results*	
	Transfers	*Consumption*	*Tranasfers*	*Consumption*
Lagged dependent level	−0.141***	−0.140***	−0.094***	−0.061***
	(−4.65)	(−5.49)	(−3.21)	(−3.55)
Trade openness	−0.008	−0.016*	0.005	0.004
	(−0.74)	(−1.73)	(0.50)	(0.44)
Capital openness	−0.192	−0.380**	−0.045	−0.012
	(−1.27)	(−2.45)	(−0.31)	(−0.12)
Left labor power (LLP)	0.067	0.134	0.116	0.196**
	(0.76)	(1.38)	(1.36)	(2.04)
LLP*Trade openness	0.001	0.001	−0.001	−0.001
	(0.41)	(1.00)	(−0.87)	(−0.86)
LLP*Capital openness	0.066**	0.075***	0.027	0.013
	(2.38)	(2.68)	(1.05)	(0.71)
Growth	−0.168***	−0.137***	—	—
	(−10.50)	(−11.53)		
Old population	0.135**	0.006	—	—
	(2.43)	(0.10)		
Unemployed	0.683***	0.008	—	—
	(3.67)	(0.42)		
Unexpected growth	—	—	−0.077***	−0.097***
			(−5.75)	(−12.14)
Automatic transfers	—	—	0.558***	—
			(6.93)	
Automatic consumption	—	—	—	0.970***
				(14.94)
Adjusted R-squared	0.4235	0.4835	0.4235	0.6735
Number of observations	350	350	350	350

[a] T-statistics are in parentheses. The results for period and country dummies are not shown.
* Significance level: < 0.10
** Significance level: < 0.05
*** Significance level: < 0.01

The results, however, turn out to be highly sensitive to the precise specification of the control variables. One of these controls is GDP growth, which Garrett explains with reference to an article by Roubini and Sachs.[23] In that article the authors argue that governments make

[23] Garrett (fn. 3, 1998), 80; Nouriel Roubini and Jeffrey D. Sachs, "Political and Economic Determinants of Budget Deficits in the Industrial Democracies," *European Economic Review* 33, no. 5 (1998).

spending decisions based on economic forecasts which rely on actual growth in the recent past. If growth turns out to be unexpectedly high, spending as a proportion of GDP will be smaller than anticipated, while spending will be higher if GDP growth is unexpectedly low.[24] They therefore define an "unexpected growth" variable, which is the difference between actual growth in a given year and average growth in the previous three years. This variable is obviously correlated with GDP growth, but it is not identical, and we have consequently substituted Roubini and Sachs's variable for Garrett's simple GDP growth variable in columns 3 and 4.

In addition, we made some refinements to the variables intended to remove nondiscretionary components of spending. In the case of transfers the relevant controls are the rate of unemployment and the size of the old-age population. These variables can be improved by taking into account the replacement rates for nonemployment that vary across time and countries. Therefore, to measure nondiscretionary transfers more accurately, we multiply the change in the size of the dependent population (that is, the proportion of unemployed and old people) by the replacement rates at any given point in time. In turn, average replacement rates can be approximated as the share of transfers in GDP relative to the share of the dependent population in the total population.[25] This composite variable is used in column 3 in place of the unemployment rate and the old-population rate.

In the case of government consumption, the number of unemployed and old people is irrelevant (as Garrett's results clearly show), but Garrett does not take into account a different nondiscretionary effect. Because costs in public services (especially wage costs) tend to increase at the same rate as in the rest of the economy, while productivity does not, a constant level of provision will cause prices of government services to rise faster than in the economy as a whole. This nondiscretionary component of government consumption can be removed by another measure, called automatic consumption, which is the share of government consumption in GDP multiplied by the rate of growth in the price deflator for government services divided by the rate of growth in the price deflator for the entire GDP.[26]

From the rise in explained variance (see Table 2), the importance of relative price changes for government consumption becomes clear.

[24] See Thomas Cusack, "Partisan Politics and Public Finance: Changes in Public Spending in the Industrialized Democracies, 1955–1989," *Public Choice* 91 (June 1997); and idem, "Partisan Politics and Fiscal Policy," *Comparative Political Studies* 32 (January 1999).

[25] Ibid.

[26] Cusack (fn. 24, 1997).

More importantly, the effects of capital market openness complet disappear once these refined controls are included. This is the ca whether consumption or transfers are examined. With respect to tra openness, one can see that the parameter on this variable is insign cant as before, but the sign on the interaction term is now actually the wrong direction. In short, Garrett's data does not support the ex tence of a relationship between globalization and welfare state spendi once more refined control variables are used. The only result that ho up is that left-labor power has a significant expansionary effect on go ernment consumption—a finding that is echoed in the more extens analysis presented below and by numerous other studies.[27]

We would like to underscore that these results do not undermi Garrett's main conclusion that globalization is compatible with a la welfare state. In fact, we agree with most of Garrett's critique of t globalization literature's predictions of broad-based retrenchment. B we also do not believe that globalization has been much of a factor the postwar expansion of the welfare state. This leaves a conspicuo gap in the general understanding of the growth of the welfare state o the past four decades. Traditional explanations that emphasize the r of the industrial working class run up against the inescapable fact th spending has skyrocketed in many countries precisely during a ti when the industrial working class has been in steep decline. This do not mean that partisanship is unimportant for the form that this e pansion has taken, as we explain below, but it does mean that, to u derstand the driving force behind the expansion, it is necessary to lo beyond standard class-power explanations.

Deindustrialization and the Labor-Market Risk Structu

Like Garrett and Rodrik we believe that exposure to risk in the lab market is a powerful determinant of peoples' preferences for state pr tection and public risk sharing, and we argue that the main sources risk are to be found in the interaction of sector-specific skills and d mestic economic processes. In particular, we suggest that the labo market dislocations associated with major shifts in the sector occupational structure have been a driving force behind the expansi

[27] For recent examples see Cusack (fn. 24); Alexander Hicks and Duane Swank, "Politics, Inst tions, and Welfare Spending in Industrialized Democracies, 1960–82," *American Political Science view* 86 (September 1992); and Evelyn Huber, Charles Ragin, and John Stephens, "Social Democ Christian Democracy, Constitutional Structure and the Welfare State," *American Journal of Socio* 99, no 3 (1993).

of the welfare state since the early 1960s. To get a sense of the numbers, in 1962 about 60 percent of the labor force in the OECD area was employed in agriculture or industry; thirty-three years later this figure had dropped to about 30 percent. As we document below, this massive sectoral shift is the outgrowth of deep forces of technological change that have coincided with progressive market saturation and shifting patterns of demand—structural-technological conditions that also characterized the industrial revolution. Given the work of Esping-Andersen,[28] Korpi,[29] Stephens,[30] and others about the relationship between the rise of industry and the early development of the welfare state, one would expect such a massive transformation of the occupational structure to be of great importance in the demand for and supply of welfare state programs.

The importance of changes in the occupational structure depends on the transferability of skills and social benefits. Transferable skills protect against market vagaries by making individuals less dependent on a single employer or on employers in a particular branch of the economy. Labor-market risks are therefore generated across the interfaces between economic sectors requiring very different types of skills. This logic is reinforced when we consider that privately provided social benefits such as health insurance and pensions also tend to be constrained by the transferability of skills. Thus, when skills are firm-specific, employers have an incentive to provide nontransferable company benefits, both as a tool of control over its workforce and as an incentive for their employees to acquire additional firm-specific skills.[31] Correspondingly, if skills are industrywide, employers in that industry have a rationale for providing benefits that are transferable across firms, but only within the industry. Although the latter depends on the ability of employers to collude in the provision of both skills and benefits, the point is that the transferability of benefits will not exceed the transferability of skills in the absence of state intervention.

The approximate correspondence between the scope of employer-sponsored insurance and the transferability of skills tells us a great deal about the sources of demand for welfare state expansion. Once a worker

[28] Gøsta Esping-Andersen, *Markets against Politics: The Social Democratic Road to Power* (Princeton: Princeton University Press, 1985); idem (fn. 7).

[29] Walter Korpi, *The Working Class in Welfare Capitalism: Work, Unions and Politics in Sweden* (London: Routledge and Kegan Paul, 1978); idem, *The Democratic Class Struggle* (London: Routledge and Kegan Paul, 1983).

[30] John D. Stephens, *The Transition from Capitalism to Socialism* (London: Macmillan, 1979).

[31] See the pathbreaking study by Isabela Mares on the role of employees in social policy formation. Mares, "Negotiated Risks: Employers' Role in Social Policy Development" (Ph.D. diss., Harvard University, 1998).

is permanently dismissed from a firm or occupation within a sector, an therefore has to either transgress a skill boundary or remain nonem ployed, both skills and benefits will be forfeited or downgraded. I some cases, then, workers are left outside employment with no or fe means of support; in other cases, workers find new jobs at substantiall reduced wages and benefits levels. Thus, workers can protect them selves against the risks of major shifts in the economic and occupationa structure only through the mediation of the state. Such protectio comes in the form of state-guaranteed health and old-age insuranc (which makes it possible to move across sectoral interfaces without los ing benefits) as well as through early retirement and certain forms o disability insurance, which facilitate a relatively painless exit from th labor market (and therefore makes it possible not to have to mov across the skill interfaces). When skills and benefits do not travel well while large numbers of people face the risks of having to make sucl "travels," we would therefore expect demand for state-sponsored com pensation and risk sharing to be high.

Such demands are not necessarily opposed by employers, as com monly assumed in the welfare state literature, and our logic highlight one of the most important reasons why. Without assurances from th state, workers will be less likely to make risky investments in nontrans ferable skills—skills that are very valuable to employers. Especially witl the transition to more knowledge-intensive forms of production, firm that rely on firm- and industry-specific skills share with their employ ees an interest in strengthening the aspects of the welfare state that re duce the riskiness for workers of making investments in specific skills Though clearly at odds with the standard perception that business al ways opposes social spending, the argument is consistent with a emerging new body of scholarship that documents the supportive an often proactive role of some groups of employers in developing an shaping the modern welfare state.[12]

Like the distinction between agriculture and industry in the latte half of the previous century, the distinction between manufacturing an services represents one of the most important economic interfaces af fecting the transferability of skills in the latter half of the twentietl century. Whereas skills within agriculture, manufacturing, or service are typically transferable to some degree, most skills acquired in eithe manufacturing or in agriculture travel very poorly to services occupa

[12] See, for example, ibid; Cathie Jo Martin, "Nature or Nurture? Sources of Firm Preference for Na tional Health Reform," *American Political Science Review* 89 (December 1995); and Peter Swensor "Arranged Alliance: Business Interests in the New Deal," *Politics and Society* 25, no. 1 (1997).

tions. Even low-skilled blue-color workers find it exceedingly hard to adjust to similarly low-skilled service sector jobs because they lack something that is vaguely referred to as social skills. In addition, as noted above, the shift in the distribution of employment between these sectors has been quite dramatic since the beginning of the 1960s. If our theoretical argument is correct, therefore, we should find at least some evidence that deindustrialization has expanded the welfare state since the early 1960s. This does not imply that no skills are transferable from industry to services or that other skill interfaces in the economy are irrelevant. Our only claim is that deindustrialization picks up one salient empirical manifestation of our theoretical logic. We are of course quite happy to concede that the use of deindustrialization as a proxy for the underlying theoretical variable—the risk of moving across skill boundaries—only establishes a lower bound for the explanatory power of our general argument.

Considering this obvious link between labor force transformations and welfare state spending, it is remarkable how little attention deindustrialization has been accorded in the study of welfare state dynamics. Not a single large-N cross-national study of the welfare state has to our knowledge focused on deindustrialization as a driving force or even included it as a control variable. Wilensky,[33] Flora and Alber,[34] and others have pointed to the importance of economic transformations, industrialization in particular, to explain the rise of the welfare state. However, the emphases in their explanations—problems associated with industrialization such as dangerous working conditions and income security for those denied access to employment[35] as well as demographic structural changes induced by the growth in overall economic well-being[36]—are very different from ours and not clearly applicable to the phenomenon of deindustrialization that we are interested in.[37]

Perhaps this omission in the literature is due to a misconception that deindustrialization is fairly uniform across countries and time and

[33] Harold L. Wilensky, *The Welfare State and Equality* (Berkeley: University of California Press, 1975).

[34] Peter Flora and Jens Alber, "Modernization, Democratization, and the Development of Welfare States in Europe," in Flora and Arnold Heidenheimer, eds., *The Development of Welfare States in Europe and America* (London: Transaction Books, 1981).

[35] Ibid., 41.

[36] See Wilensky (fn. 33), 47.

[37] There are passages in Harold L. Wilensky and Charles N. Lebaux's *Industrial Society and Social Welfare* that concern the problem of relocating and retraining workers made redundant in the process of industrialization in the U.S. This logic is clearly in line with our argument. We thank an anonymous reader for pointing this out. Wilensky and Lebaux, *Industrial Society and Social Welfare: The Impact of Industrialization on the Supply and Organization of Social Welfare Services in the United States* (New York: Russell Sage Foundation, 1958).

therefore cannot explain cross-national and temporal variance in the speed of welfare state expansion. In fact, however, deindustrialization varies greatly in both time and space. For example, in an early industrializing country like the United States, industrial employment as a percentage of the adult population never reached the same levels as in a late industrializer like Sweden—22 versus 31 percent at their peaks. The decline was correspondingly much slower and modest in the U.S., falling by 3 percent between 1962 and 1993 compared to 13 percent in Sweden. In general, the amplitude of the swings in the sectoral employment structure is much greater in late than in early industrializers, which accounts for most of the variance in the speed of deindustrialization. As we show below, the best predictor of deindustrialization is simply previous levels of industrialization.

Both the magnitudes of the sectoral shifts in employment and the cross-national differences are magnified by the decline of agriculture. Although we usually associate agricultural decline with the rise of industry, the two processes started to move in phase in the early to mid-1960s, particularly in countries that industrialized late. Agricultural decline is due to the same forces of structural-technological change, which is explored below. Hence, when we talk about deindustrialization in the following, we have in mind this secular, long-term, and structurally driven process of labor shedding in *both* agriculture and industry beginning in the early 1960s.

While we maintain that deindustrialization is a crucial (and neglected) source of welfare state expansion, we are not implying that political and institutional factors are unimportant. As Garrett emphasizes in his work, the welfare state is a mechanism for redistribution as well as risk sharing. We would therefore expect partisan governments and organized interests to shape social policies in order to benefit the distributive interests of their own constituencies. As argued well by Garrett, where unions are strong and centrally organized and where left governments have been dominant, the welfare state can be expected to assume a more redistributive form.[38] Likewise, redistribution is affected by the location of the median voter insofar as political parties adopt policies that will appeal to the median voter.[39] The lower the income of

[38] Garrett (fn. 3).

[39] The attraction of the median voter depends on the electoral system. In two-party systems the mechanism is vote maximization, which leads to centrist appeals. See Anthony Downs, *An Economic Theory of Democracy* (New York: Harper and Row, 1957), chaps. 7–8; and Gary W. Cox, "Centripetal and Centrifugal Incentives in Electoral Systems," *American Journal of Political Science* 34, no. 4 (1990). In multiparty systems the mechanism is office maximization through which governing coalitions come to recognize the importance of appealing to the median voter. See Michael Laver and Norman

the median voter and the more exposed to risk, the greater the pressure for redistributive policies.[40] Since low-income workers in tenuous labor-market positions are less likely to vote than better educated and higher-income people,[41] an indirect measure of the median voter location is the level of participation in national elections.

The explanatory salience of these political variables depends on the extent to which we look at spending categories that have a redistributive effect. Aggregate levels of transfers are not necessarily higher under left than under right governments insofar as such transfers can be used to address labor-market risks without affecting income or status differentials.[42] By contrast, government service provision is inherently redistributive because it offers people equal access to services, such as education, health care, and housing, which are paid for through taxation. In addition, egalitarianism and public sector expansion are causally related because earnings compression undermines the growth of low-productivity and price-sensitive private service sector jobs, thereby putting pressure on the government to provide "compensating" jobs in the public sector.[43] So while deindustrialization everywhere propels the growth of welfare state spending, whether in the form of government transfers or consumption, we expect the distributive aspects of the rising service economy and the private-public sector mix of employment to vary according to political parameters.

FINDINGS

We use an error correction model of the type introduced in Table 2, with changes in government transfers and civilian government consumption as the dependent variables. The model has the following form:

$$\Delta Y_{i,t} = \alpha + \beta_1 \; Y_{i,t-1} + \Sigma\beta^j \cdot X^j_{i,t-1} + \Sigma\beta^j_\Delta \cdot \Delta X^j_{i,t} + \varepsilon_t,$$

where Y is a spending variable and X is an independent variable. The subscripts i and t refer to the particular country and time period re-

Schofield, *Multiparty Government: The Politics of Coalition in Europe* (Oxford: Oxford University Press, 1991).

[40] Alan Meltzer, and Scott Richard, "A Rational Theory of the Size of the Government," *Journal of Political Economy* 8, no. 9 (1981). See also Hicks and Swank (fn. 27).

[41] Arend Lijphart, "Unequal Participation: Democracy's Unresolved Dilemma—Presidential Address," *American Political Science Review* 91, no. 1 (1997).

[42] Esping-Andersen (fn. 7).

[43] See Gøsta Esping-Andersen, *Changing Classes: Stratification and Mobility in Postindustrial Societies* (London: Sage, 1993); idem, *Social Foundations of Postindustrial Societies* (Oxford: Oxford University Press, 1999); Iversen and Wren (fn. 8); and Andrew Glyn, "Low Pay and the Volume of Work" (Manuscript, Corpus Christi College, Oxford University, 1997).

spectively, while the superscript *j* refers to the particular independent variable. Δ is the first difference operator.

This model has a number of useful properties, and it is consistent with recommendations on specifications that are capable of capturing both long- and short-term dynamics in a pooled time-series-cross-section context.[44] First, the parameter for the lagged dependent level variable (β_1) provides an easy check on equilibrium properties. β_1 should be between -1 and 0 to ensure that the incremental effects of a shock to any exogenous variable are progressively reduced over time, causing spending to converge to a long-term equilibrium. For readers more familiar with models that use the *level* of spending on the left-hand side, the current model can be reformulated into such a model by simply adding $Y_{i,\,t-1}$ on both sides of the equal sign. This yields $Y_{t,\,i} = \alpha + (1 + \beta_1)\ Y_{i,\,t-1} + \ldots$, where $(1 + \beta_1)$ is the new parameter for the lagged dependent level variable. There is a small advantage to using the error correction formulation, however, because the model yields estimates of R^2 that are more informative of the variance explained by the independent variables of interest.[45] Otherwise the choice to use either levels or first differences does not affect the results.

The other useful feature of the present model is that it enables us to separate out permanent and transitory effects of any independent variable. Although not intuitively obvious, it can be shown that the parameter for a lagged independent *level* variable, $X_{t-1,}$ is a measure of the permanent (or lasting) effect of a one-off change in that variable, while the parameter for a *change* variable, $\Delta X_{t,}$ is a measure of the transitory (or passing) effect of a one-off change in that variable.[46] The long-term permanent effect of an independent variable can be calculated by dividing the parameter for the lagged level of that variable by minus the parameter for the lagged dependent level variable: $\beta^j/{-\beta_1}$ (assuming that β_1 is between 0 and -1). If a variable exhibits *only* transitory effects, that is, if only the parameter for its first difference is different from zero, spending will eventually revert back to its original level unless the independent variable changes continuously (assuming again that β_1 is between 0 and -1). Since all the theoretical variables are defined as

[44] See Nathaniel Beck and Jonathan N. Katz, "Nuisance vs. Substance: Specifying and Estimating Time-Series-Cross-Section Models," in John R. Freeman, ed., *Political Analysis*, vol. 6 (Ann Arbor: University of Michigan 1996), 11.

[45] The reason is that in the model using levels of spending on the left-hand side, much of the variance will be accounted for by the lagged dependent variable, showing simply that current spending depends on past spending.

[46] See Nathaniel Beck, "Comparing Dynamic Specifications: The Case of Presidential Approval," in James Stimson, ed., *Political Analysis*, vol. 2 (Ann Arbor: University of Michigan Press, 1992).

proportions (either of GDP or of the working-age population), they cannot grow (or fall) without limit and will therefore have no lasting effects on spending unless the parameters for the their lagged levels are nonzero.[47] Hence, the parameters for the change variables are of interest only if we care about the specific time dynamics of an independent variable. To keep the results simple, we have therefore only included first differences for those independent variables that are of particular theoretical interest.

We use fairly much the same set of explanatory variables for both transfer spending and civilian government consumption outlays. The exact variable definitions and data sources are summarized in the Appendix. The only difference between the two specifications is the "autonomous" spending term in each equation. In the equation for transfers, this item is based on the prevailing replacement rates of the program (at time t–1) times the first difference in the size of the clientele for such programs. In the equation for government consumption, the autonomous spending term is a function of the prevailing level of spending (at time t–1) times the rate of change in the relative prices confronting government. As discussed above, in both instances the argument is that there are nondiscretionary elements to spending that need to be eliminated in any well-specified model.[48]

In addition to the lagged level of the spending component, there are four sets of variables in each specification. First, there is a set of variables meant to detect whether international or domestic economic sources are driving spending. On the international side, we have included measures of trade openness as well as capital mobility. On the domestic side, we have introduced measures for deindustrialization and for a variety of political variables such as the level of electoral turnout, the left-right partisan composition of the government, and a measure of the relative strength of labor within the industrial relations system. The remaining control variables have already been introduced in the discussion of Garrett's results.

Deindustrialization is defined as 100 minus the sum of manufacturing and agricultural employment as a percentage of the working-age

[47] This does not *have* to be the case. One of the control variables, unexpected GDP growth, can in principal rise indefinitely.

[48] The basic results we report below for the deindustrialization and globalization variables are robust to a large number of alternative model specifications. Thus, we have tried to group the data into time intervals using different periodizations; we have run the regression with and without the political variables, with and without the change terms; and we have tried to exclude one country at a time. In all cases the results for deindustrialization are strong and statistically significant, whereas for the globalization variables, they are weak and statistically insignificant, with the exceptions noted in the text.

population. The base of 100 is somewhat arbitrary. For example, the peak of employment in agriculture and manufacturing, a number that varies across countries, could have been used as the base instead. However, the statistical analysis is insensitive to the choice of base due to the inclusion of a full set of country dummies.[49] If each country has a unique base, it simply alters the nationally specific intercepts that can take on any value. Also, it should be emphasized that because deindustrialization is defined as a proportion of the working-age population it is not the case that public employment, which is closely related to government consumption, will alter the denominator. The only way that government consumption, or any other form of government spending, can affect the deindustrialization measure is if spending is causally related to the number of people employed in agriculture or industry—a possibility we will consider in the next section.

The equations have been estimated using a pooled data set with fifteen countries and a temporal domain ranging from 1961 up to and including 1993, a period of thirty-three years. Tests for heteroskedasticity in both pooled regressions suggested the need to correct for this problem and so we employed Beck and Katz's method for deriving panel corrected standard errors.[50] The results for two simple additive models, one for each category of spending, are presented in Table 3. Separate runs using robust regression techniques (not shown) yield almost identical results, so our findings are not driven by outliers.[51]

First, note that none of the globalization variables registers a statistically significant permanent impact on government transfers, only a small transitory effect of trade openness. For consumption, trade and capital market openness both exhibit small significant effects, but for capital mobility the effect is entirely transitory, while for trade it goes in the opposite direction of that predicted by the openness argument. It is conceivable that the negative effect for trade reflects its differential welfare effects. Thus, while growing exposure to competition from low-wage countries raises the uncertainty for those already at high risk,[52]

[49] An F-test indicates that the country dummies belong in the model.

[50] Nathaniel Beck and Jonathan N. Katz, "What to Do (and Not to Do) with Time-Series-Cross-Section Data," *American Political Science Review* 89, no. 3 (1995).

[51] Specifically, we used Roy E. Welch's "bounded-influence technique," which weights individual cases if they exceed a size-adjusted cut-off point. Welch, "Regression Sensitivity Analysis and Bounded-Influence Estimation," in Jan Kmenta and James B. Ramsey, eds., *Evaluation of Econometric Models* (New York: Academic Press, 1980), 164–66. See also J. Fox, *Regression Diagnostics* (London: Sage Publications 1991), 31.

[52] Andrian Wood, *North-South Trade, Employment and Inequality: Changing Fortunes in a Skill-Driven World* (Oxford: Oxford University Press, 1994); Edward E. Leamer, *Sources of International*

TABLE 3
REGRESSION RESULTS FOR GOVERNMENT SPENDING[a]

		Transfers	Consumption
	Lagged spending level (Y_{t-1})	−0.067***	−0.051***
		(−3.17)	(−3.78)
Globalization variables	Trade openness	−0.005	−0.004*
		(−1.21)	(−1.72)
	Δ Trade openness	0.018	−0.005
		(2.09)**	(−0.99)
	Capital openness	0.010	−0.007
		(0.35)	(−0.39)
	Δ Capital openness	0.016	−0.069**
		(0.29)	(−2.07)
Deindustrialization variables	Deindustrialization	0.044***	0.031***
		(3.10)	(3.18)
	Δ Deindustrialization	0.142***	0.090***
		(3.69)	(4.08)
Political variables	Left government CoG	−0.062	0.090***
		(−1.24)	(2.67)
	Δ Left government CoG	0.041	0.049
		(0.62)	(1.19)
	Electoral participation	−0.005	0.012***
		(−0.69)	(2.67)
	Strength of labor	0.078	0.898***
		(0.09)	(2.88)
Controls	Unexpected growth	−0.077***	−0.092***
		(6.43)	(14.59)
	Automatic transfers	0.845***	
		(9.53)	
	Automatic consumption		0.971***
			(15.87)
	Adjusted R-squared	0.47	0.63
	Number of observations	495	495

[a] T-scores are in parentheses. The results for country dummies are not shown.
* Significance level: <0.10
** Significance level: <0.05
*** Significance level: <0.01

trade may well improve welfare for all others.[53] Whatever the explanation, the magnitude of the effect is small. Thus, for each percentage point that the foreign sector grows, the long-term equilibrium level of civilian government consumption declines by only 0.07 percent.

Compare these results with those for deindustrialization. For each percent decline in employment in the traditional sectors, the long-term target equilibrium for social transfer spending increases by approximately 0.4 percent. The corresponding effect for government consumption is 0.6 percent, while the short-term impact is the elevation of the actual spending level by 1 percent for every percent decrease in employment in the traditional sectors. In other words, a standard deviation change in deindustrialization is associated with roughly one-half of a standard deviation change in spending, which implies that about half of the variance in spending is explained by the deindustrialization variable. All effects of deindustrialization are statistically significant at a 0.01 level or better. From these results it seems justified to conclude that the effects of the domestic economic variables carry far greater weight than globalization in shaping government spending.[54]

Another feature of the findings deserves emphasis: the effect of deindustrialization persists over time. Apparently spending gets locked in by organizational and institutional factors that are exogenous to our model. As argued by Pierson, spending itself creates political clienteles that will press for further spending and resist attempts at retrenchment.[55] Hence, even though the process of deindustrialization is the causal agent in the expansion of the welfare state, the disappearance of this causal agent will not necessarily lead to retrenchment but will "merely" retard further expansion. However, the character of the political game over welfare policies is likely to change when compromises

Comparative Advantage: Theory and Evidence (Cambridge: MIT Press, 1984); idem, "Trade Economists View of U.S. Wages and 'Globalization'" (Paper presented at the Political Economy of European Integration Study Group Meeting, University of California, Berkeley, January 1996).

[53] Rodrik (fn. 2), chap. 4.

[54] We tested whether the effects of employment losses in agriculture differ from those in manufacturing by breaking the deindustrialization variable into two component variables: one variable defined as 100 minus manufacturing employment as a percent of the working-age population, and another defined as 100 minus agricultural employment as a percent of the working-age population. The results for both variables are similar to those for the deindustrialization variable, but when both variables are entered simultaneously, the manufacturing variable has a somewhat stronger effect (0.045 versus 0.035 for transfers, and 0.035 versus 0.022 for consumption). One reviewer suggested that a possible reason for this pattern can be found in social mobility studies, which show that a greater decline in agriculture occurs through intergenerational mobility than in manufacturing.

[55] Paul Pierson, *Dismantling the Welfare State: Reagan, Thatcher and the Politics of Retrenchment* (New York: Cambridge University Press, 1994); idem, "The New Politics of the Welfare State," *World Politics* 48 (January 1996).

involving overall expansion are no longer feasible—a conjecture that deserves further exploration, considering that the process of deindustrialization is coming to a halt in many countries.

None of the political terms register any impact on transfers. As discussed previously, the level of transfer payments is not necessarily a contentious partisan issue, unlike the distributive composition of such payments. As Esping-Andersen notes, "There is no reason to expect that expenditure commitments, as such, should be related to left-party power."[56] Right as well as left governments, exposed to the pressures of democratic politics, recognize the need to address the risks that people encounter in the labor market; these risks are largely captured by the deindustrialization variable. Where they obviously differ is in terms of whose interests in the electorate are accorded more or less attention, a distributive issue to which the aggregate level of transfers does not speak.

In this respect the logic of government consumption is very different because public provision of services directly reduces inequalities in peoples' access to basic services such as education and health care and because public employment is used by left governments to support egalitarian wage policies.[57] Unsurprisingly, therefore, all of the political variables turn out to affect civilian government consumption in the predicted direction. Thus each percentage increase in the electoral participation rate raises the target level of spending by about 0.15 percent. Likewise, a typical left government spends about 2 percent more than a typical right government if we look at the long run.[58] The strength of labor in the industrial relations system also has an upward effect on spending, as expected.

These results, however, disguise a richer causal story that cannot be captured by simple additive models like those presented in Table 3. As argued by Garrett and others, the effects of forces that create labor-market risks are conditional on political and institutional factors, and our argument implies several causal pathways for the relationship between deindustrialization and spending. In order to pinpoint these mechanisms and to bring out the interaction effects that are implied by the theoretical argument, we have conducted a causal path analysis. The results from this analysis are summarized in Figure 2.

The effect of deindustrialization runs along two different paths. First, deindustrialization raises the generosity of transfer payments as

[56] Esping-Andersen (fn. 7), 115.

[57] Ibid.; Iversen and Wren (fn. 8).

[58] A typical left government is defined here as one that is 1 standard deviation to the left of the mean on the partisan government variable. A typical right government is defined similarly.
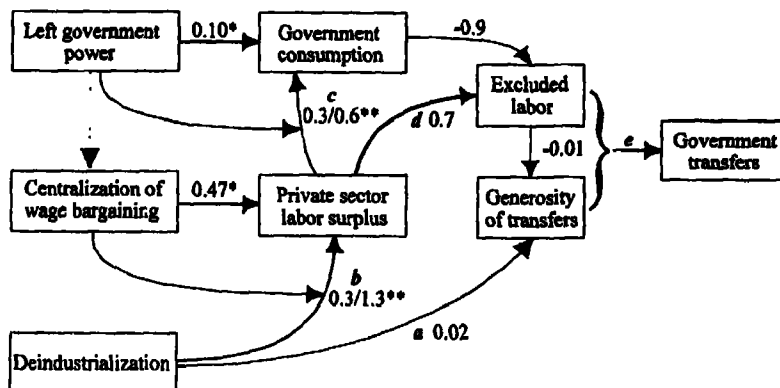
FIGURE 2
THE CAUSAL MECHANISMS LINKING DEINDUSTRIALIZATION TO SPENDING[a]

[a] All parameters are long-term, permanent effects; all are significant at a .05 level or better. Variables are defined in the text and in the appendix.

KEY:
a: Deindustrialization creates electoral pressures to ensure against labor-market risks (higher "generosity of transfers").
b: Deindustrialization leads to redundancies when private service sector employment growth is constrained (which is greater when wage bargaining is centralized).
c: Employment problems generate electoral pressures to expand government services (which are accommodated primarily under left governments).
d: Redundant labor that is not being employed in either private or public services will increase the number of people reliant on transfers ("excluded labor").
e: Government transfers are equal to the product of "excluded labor" and "generosity of transfers."
*: Long-term effect when deindustrialization (alternatively: private sector labor surplus) is at its mean value; **: effect of deindustrialization (alternatively: private sector labor surplus) when the conditioning variable is 1 standard deviation below or above its mean.

governments respond to electoral pressures for insurance against labor-market risks (path *a*). Generosity is measured here as the ratio of transfers to GDP over the ratio of the nonworking to the total population. Thus, if transfers as a proportion of the total size of the economy rise faster than the share of the nonworking population, generosity will increase. As before, we use a pooled error correction model with changes in generosity on the left-hand side, and the lag of the level of generosity plus all the variables in the second column of Table 3 on the right-hand side. The coefficient next to the causal arrow measures the permanent long-term effect of deindustrialization. In substantive

terms, the estimated coefficient suggests that a 1 percent loss of employment in the traditional sectors will raise the level of generosity by a similar amount, adjusting for all indirect effects. In turn, higher generosity obviously implies higher government transfers, so deindustrialization is directly related to spending via generosity.

The second causal path goes through the employment effects of deindustrialization, which are conditioned by labor-market institutions and partisan politics. Insofar as redundant labor in the traditional sectors are not being picked up by new employment in private services, (path *b*) governments are likely to face electoral pressures to create employment by expanding public provision of services (path *c*). This effect, however, is conditioned by two political-institutional variables. First, we know from past studies that the capacity of the private service sector to generate employment is negatively related to wage compression.[59] The likely reason for this negative relationship is that productivity grows at a much slower rate in most services than in most manufacturing. Hence, in labor-intensive services, which include many personal and social services, a tightly coupled and compressed wage structure will result in rising relative prices and therefore in a slower rate of new job creation. Because centralized wage bargaining is closely related to wage compression,[60] this phenomenon is observed primarily in northern Europe where bargaining is highly centralized. And it shows up clearly in our results as an interaction effect between deindustrialization and centralization of bargaining.

More specifically, in liberal market economies with decentralized bargaining, such as Canada or the United States, most or all employment losses in the traditional sectors are picked up by growth in private service employment. Surplus labor in the private sector, measured as the difference between deindustrailization and changes in private service sector employment, is consequently small or absent. By contrast, when bargaining is centralized, as in Germany or Sweden, deindustrialization produces large numbers of workers who cannot find employment in private services. In terms of our regression analysis, which includes controls for labor-force participation rates and economic

[59] See Eileen Appelbaum and Ronald Schettkat, "Employment and Productivity in Industrialized Countries," *International Labor Review* 134, nos. 4–5 (1995); and Iversen and Wren (fn. 8).

[60] Robert Rowthorn, "Corporatism and Labour Market Performance," in Jukka Pekkarinen, Matti Pohjola, and Rowthorn, eds., *Social Corporatism* (Oxford: Clarendon Press, 1992); Michael Wallerstein, "Wage-Setting Institutions and Pay Inequality in Advanced Industrial Societies," *American Journal of Political Science* 43 (July 1999); David Rueda and Jonas Pontusson, "Wage Inequality and Varieties of Capitalism" (Working paper, Institute for European Studies, Cornell University, 1998).

growth, when centralization is one standard deviation below the mean, a 1 percent drop in manufacturing employment results in a 0.3 percent increase in the private sector employment surplus. When centralization is one standard deviation above the mean, this effect is about four times greater.[61]

There is a close relationship between left partisan control over government power and centralization of wage bargaining (r = 0.8). Although this relationship has emerged only through a complex historical process without a clear-cut causal order, it is among the liberal market economies—those dominated by flexible labor markets and liberal governments—where governments are faced with the fewest problems of private sector labor redundancies. Instead, these countries "pay" for deindustrialization in the form of higher wage inequality, partly cushioned by the increased generosity of transfer payments (path *a*). Other countries have dealt with the employment pressures from deindustrialization in ways that also depend on the partisan orientation of the government. Specifically, wherever the left holds strong positions in government, service employment is expanded by increasing the direct provision of government services, that is, through a rise in government consumption (path *c*). In substantive terms, when left partisanship is one standard deviation above the mean, a one standard deviation increase in the private sector labor surplus is associated with a commensurable increase in government consumption. When considering countries with centralized bargaining systems, right governments (of a mostly Christian democratic bent) shy away from expanding public service production. The increase in consumption in these cases is correspondingly less than half the increase in the labor surplus of the private sector.

Finally, when redundant labor is not absorbed into public service employment, the private sector labor surplus will register in the form of early retirees, disability pensioners, and unemployed workers (labeled *excluded labor* in Figure 2) who have to be cared for through a variety of cash benefits arrangements (path *d*). This leads to an increase in government transfers, the size of which depends on the generosity of these transfers ("transfers" is simply "generosity" times "excluded labor"). The increase is not directly proportional to the increase in the excluded labor force because the latter has a dilution effect on the generosity of transfers, presumably because governments will seek to limit the bud-

[61] Both figures are slightly exaggerated because there are small negative *indirect* effects of deindustrialization (through growth and other variables) that reduce the net impact of deindustrialization. Yet these indirect effects do not affect the *gap* in the effect of deindustrialization between countries with centralized and decentralized bargaining structures, which is what is of interest here.

getary pressures by cutting other welfare programs. Still, the net effect is a significant increase in total government transfers.

In combination, the results from these regression analyses strongly support the deindustrialization argument. Thus, as deindustrialization increases, both government transfers and government consumption rise, but the latter effect is particularly strong when wage bargaining is centralized and when governments are dominated by left parties. Under right governments, most of the effect comes through increases in transfers. The reason is that right governments will not compensate for private sector labor redundancies through public service production, using instead transfer schemes, such as early retirement, to facilitate labor-market exit. Since centralized wage bargaining reduces the absorptive capacity of workers who are made redundant in the traditional sectors, the effect on transfers is particularly strong when bargaining is centralized.

THE SOURCES OF DEINDUSTRIALIZATION

Our results strongly suggest that deindustrialization, not trade or capital market openness, is the driving force behind the expansion of government spending on both transfers and services. Nevertheless, it could be objected that deindustrialization may itself be a consequence of trade and financial openness or that it is caused by, not causing, government spending. Even though either one of these possibilities is interesting in its own right, both would obviously significantly alter the picture we have presented of the relationship between deindustrialization and spending. In order to complete our argument, we therefore have to show that deindustrialization is largely driven by domestic factors other than spending itself.

Economists are divided on the question of whether trade causes employment losses in the traditional sectors. On one side of the debate, reflecting not only a particular economic theory but also a generally popular view (the "giant sucking sound"), is the idea that the sources of deindustrilization in the West during recent decades lay squarely in the competitive pressures emanating from Third World producers.[62] From this perspective, changes in the North-South trade have been estimated to account on average for 50 percent of the reduction in manufacturing that occurred between 1970 and 1990.[63] In addition, it can be argued

[62] See, for example, Wood (fn. 52); Steven S. Saeger, "Globalization and Economic Structure in the OECD" (Ph.D. diss., Harvard University, 1996); and idem, "Globalization and Deindustrialization: Myth and Reality in the OECD," *Weltwirtschaftliches Archiv* 133, no. 4 (1997).

[63] Saeger (fn. 62, 1997), 604.
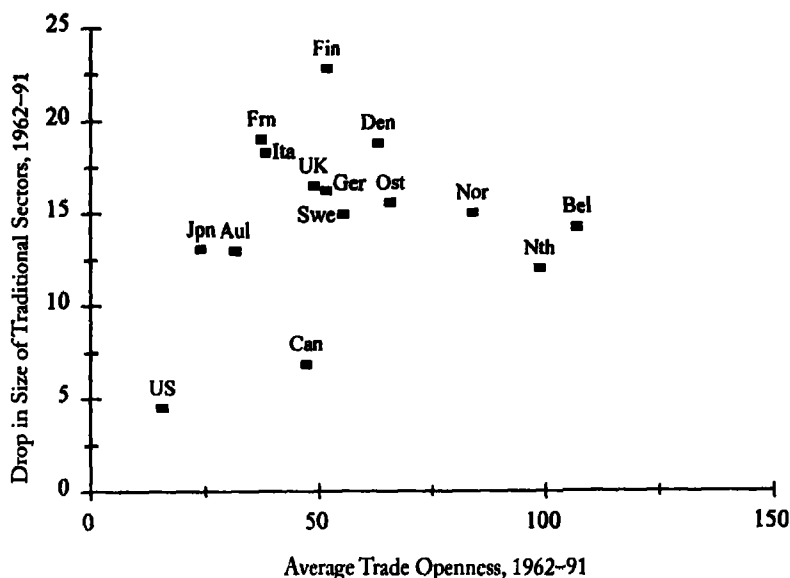
FIGURE 3
TRADE OPENNESS AND LOSSES IN TRADITIONAL SECTORS[a]

[a] Trade openness is exports plus imports as a percentage of GDP; drop in the share of traditional sectors is measured as the change in the percentage of the working-age population engaged in agriculture or industry between 1962 and 1991.

that the removal of restrictions on capital makes it increasingly easier for businesses to relocate production facilities to countries with lower wage costs and that this in turn diminishes the demand for labor within the industrial sectors of the advanced market economies.[64]

The alternative school, while not denying that trade has played a role in deindustrialization, sees the principal causes as residing in domestic sources.[65] Among these are a change in preference patterns away from manufactured goods and toward services, high productivity growth in the face of inelastic demand, as well as the associated changes in in-

[64] Wolfgang Streeck, "German Capitalism: Does It Exist? Can It Survive?" in Colin Crouch and Streeck, eds., *Political Economy and Modern Capitalism: Mapping Convergence and Diversity* (London: Sage, 1997).

[65] Robert Rowthorn and Ramana Ramaswamy, "Deindustrialization: Causes and Implications" (International Monetary Fund Working Paper, no. 42, Washington, D.C., 1997); idem, "Growth, Trade, and Deindustrialization, *International Monetary Fund Staff Papers* 46 (March 1999); Paul Krugman, *Pop Internationalism* (Cambridge: MIT Press, 1996).

vestment in new productive capacity.[66] North-South trade accounts for at most one-sixth of the loss in manufacturing employment in these studies.[67]

Furthermore, it may indeed be the case that the welfare state is itself responsible for the decline in employment in the traditional sectors. As Bacon and Eltis have argued, both the costs posed by taxation as well as the generosity of the modern welfare state, including the opportunity to work for at least equivalent if not higher wages in the public sector, have had a tremendous negative effect on industrial employment.[68] The notion of efficiency losses from distortionary taxation is in fact a quite common assumption in economic models of the welfare state.[69] It is of course also a view that is popular with political parties and governments of a neoliberal bent.

Figure 3 provides some descriptive evidence on the question of whether trade causes deindustrialization. It plots the loss of employment in the traditional sectors from 1962 through 1991 against the average trade openness for the same period. There is little hint of any relationship. Indeed the correlation between the two series is about 0.17.

Alternatively, if one were to adopt the hypothesis that deindustrialization has more to do with internal processes, processes of productivity gain and shifting tastes, then one would expect that a process of convergence has been under way. Thus, early industrializers, who had pretty much gone through this transformation by the beginning of this period, would have suffered the least loss of employment in the traditional sectors, while late industrializers would have experienced more rapid decline. As Figure 4 demonstrates, there seems to be a fair amount of support for this position. The correlation between employment intensity in the traditional sectors in the year 1962 and the loss of employment in these sectors over the three succeeding decades is about 0.85. Thus, the United States, which had the smallest traditional sectors (about 24 percent), experienced the smallest loss (less than 5 percent). Finland, lagging well behind the United States and having nearly

[66] Rowthorn and Ramaswamy (fn. 65, 1999), 19.

[67] We recognize that the debate over trade with LDC countries is broader than the question of whether it has led to widespread deindustrialization. It also involves issues such as the relative performance of particular manufacturing industries, relative prices, and above all, wage inequality. For a good introduction to the debate see the special issue of *Journal of Economic Perspectives* 9 (Special issue, 1995). For our purposes, however, the key issue is whether or not our main independent variable, deindustrialization, is in large measure the result of competition from low-wage countries.

[68] Robert Bacon and Walter Eltis, *Britain's Economic Problem: Too Few Producers* (London: Macmillan, 1976).

[69] See, for example, Meltzer and Richard (fn. 40); and Alberto Alesina and Roberto Perotti, "The Welfare State and Competitiveness," *American Economic Review* 87 (December 1997).
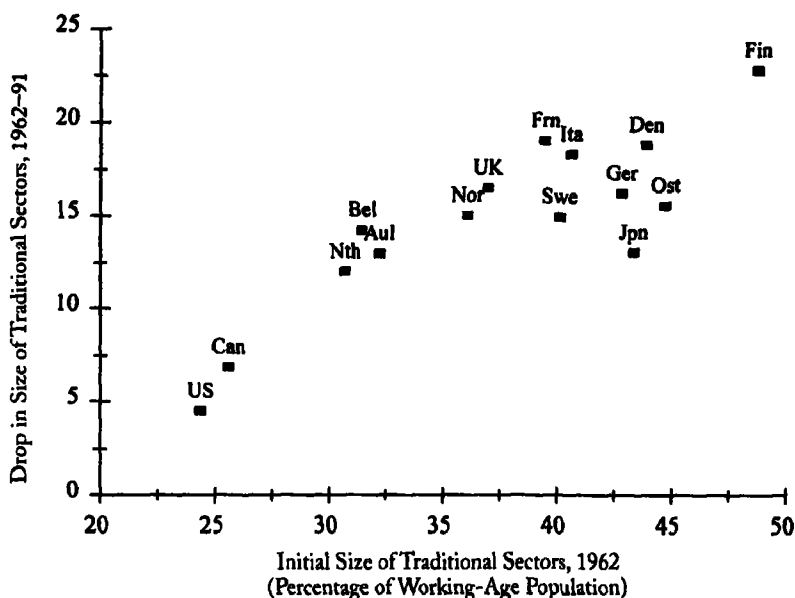
FIGURE 4
INITIAL SIZE AND LOSSES IN TRADITIONAL SECTORS

50 percent of its working-age population engaged in the traditional sectors, experienced the largest loss in the sample of fifteen countries, well over 20 percent.

But descriptive and indirect evidence of this nature can sometimes be misleading. We have therefore estimated a pooled cross-section-time-series model that uses the change in the log of the number of people employed in manufacturing and agriculture as a share of the working-age population as the dependent variable (see Table 4).[70] This setup is standard in the existing literature, except that we have included agricultural employment on the left-hand side to make the results speak directly to our deindustrialization variable.[71] However, the results are very similar if we focus exclusively on manufacturing employment.

[70] As in the previous analysis, problems of heteroskedasticity led us to employ Beck and Katz's method for deriving panel corrected standard errors. Beck and Katz (fn. 50).

[71] We could have used change in the log of the deindustrialization variable on the left-hand side in Table 4 without affecting the results (except for switching the signs of parameters, of course). We chose the current setup to stay as close as possible to the analysis by Rowthorn and Ramaswamy (fn. 65).

TABLE 4
REGRESSION RESULTS FOR INDUSTRIALIZATION[a]

Domestic Variables		*International Variables*	
[Lagged level]	−0.113***	Capital openness	0.001
	(−5.27)		(0.30)
Productivity growth	−0.353***	Δ Capital openness	0.001
	(9.09)		(0.50)
Income	0.523**	OECD trade balance	0.002***
	(2.18)		(3.47)
Income squared	−0.30**	Δ OECD trade balance	0.004***
	(2.24)		(4.43)
Growth in income	0.585***	OPEC trade balance	−0.004*
	(8.50)		(−1.96)
Capital formation	0.032***	Δ OPEC trade balance	−0.003
	(5.53)		(−1.35)
Government transfers	−0.001	LDC trade balance	−0.003**
	(−0.99)		(−2.12)
Government consumption	0.001	Δ LDC trade balance	−0.002
	(0.30)		(−1.19)
Increase in explained var.	35%		5%
Adjusted R-squared			0.52
Number of observations			378

[a] T-scores in parentheses. The increase in explained variance is the change in R-squared when the set of endogenous and exogenous variables are added to a model where these variables are excluded. The results for country dummies are not shown.
* Significance level: <0.10
** Significance level: <0.05
*** Significance level: <0.01

The analysis includes fourteen OECD countries for which we have complete data in the period from 1964 through 1990.[72]

For presentational ease Table 4 divides the independent variables in a group of "domestic" variables and a group of "international" variables. Following the existing economic literature we include among the domestic-structural variables: (1) a measure of productivity growth, (2) the log of income per capita and the square of this variable to capture

[72] The countries include: Austria, Belgium, Canada, Denmark, Finland, France, Germany, Italy, Japan, Netherlands, Norway, Sweden, the United Kingdom, and the United States. Missing data problems precluded adding Switzerland. The time frame is the maximum possible given the availability of data.

changing consumption preferences, (3) the growth in per capita income as a measure of demand effects, (4) gross capital formation as a share of GDP, and (5) the two spending variables. For the exogenous variables we have included (1) the balance of trade with OECD, with OPEC, and with less-developed countries (LDCs); and (2) the capital market openness variable used above.

The productivity measure is meant to capture the tendency of firms to shed workers as productivity increases. Note that there is some theoretical ambiguity with respect to the impact of this variable. While faster productivity growth makes goods relatively cheaper and therefore boosts demand, less labor is required to produce the same amount of output. The net effect on employment depends on the price and income elasticity of demand as well as on real wage changes. Research, however, has shown that the labor saving effect tends to dominate the demand effect in the period of interest.[73] For the income terms, the expectation is that the parameter on the first term will be positive and that on the second term will be negative, signifying that as income passes beyond a certain level, the relative demand for goods in both the agricultural and manufacturing sectors will begin to decline. The effects of capital formation and growth in income are expected to be positive since both will boost production and demand for labor.[74]

The results are generally very supportive of our argument. Deindustrialization is almost exclusively driven by domestic factors other than the welfare state. Technological progress, demand conditions, and shifting consumption patterns cause employment in industry and agriculture to decline. Evidence does not show that government spending has crowded out employment in the traditional sectors. Given our previous findings, every indication is that the causal arrow goes in the opposite direction. Trade also does not appear to be an important source of deindustrialization. A negative trade balance with other industrialized countries (and the first difference in that trade balance) does hurt industrial employment. The effect is substantively small, however, and cannot have been a major cause of deindustrialization across the OECD area for the simple reason that intra-OECD trade remains relatively balanced over time.

The crucial question with respect to trade is whether growing trade with less-developed countries has priced out a substantial number of

[73] See Appelbaum and Schettkat (fn. 59); and idem, "The End of Full Employment? On Economic Development in Industrialized Countries," *Intereconomics* 29 (May–June 1994).

[74] Investment is measured as a percentage share of GDP. It is taken from the Robert Summers and Alan Heston, *Penn World Tables*, version 5.5, datafile (Cambridge: National Bureau of Economic Research, 1993).

workers in agriculture and industry in the advanced countries. We find no evidence to that effect. The coefficients on the lagged levels of the trade balances with OPEC countries and with Third World countries are both negative and statistically significant, while both of the coefficients on the first differences are statistically insignificant. Note, that these results, which suggest that positive trade balances with the OPEC and Third World countries lower employment while negative balances promote employment, are not the consequence of multicollinearity. Also, their effects do not change in substantive terms when we use alternative specifications of the model. We have run a large number of regressions using a variety of combinations of trade balances and import penetration measures, and the results are *all* contrary to the hypothesis that trade leads to deindustrialization. In fact, the results in Table 4 are the strongest we have been able to produce in *support* of the popular perception.[75] The same is true for the capital-market-openness variable which consistently fails to produce effects that are statistically distinguishable from zero.

Though surprising given popular views, our results essentially replicate those in an OECD study by Rowthorn and Ramaswamy,[76] even though our data and model specification are somewhat different. Like Rowthorn and Ramaswamy we find that deindustrialization is driven primarily by deep economic processes that are unrelated to either openness or spending. Productivity growth in the traditional sectors leads to a loss in employment, whereas rising demand through growing investment or incomes has a positive effect. Consistent with Engel's law, the results also indicate that demand for agricultural and manufacturing first rises with income and then falls at higher levels thereby eventually diminishing the level of traditional employment. We conclude from this analysis that our argument and results for spending are robust to the charges that deindustrialization is a mediating variable and that its association with spending is a result of reversed causality.

Conclusion

The domestic effects of the international economy has been increasingly emphasized in political-economic theory as well as popular accounts. There is no denying that international trade and financial liberalization have heightened interdependence among states and

[75] The results of alternative specifications are available from the authors upon request.
[76] Rowthorn and Ramaswamy (fn. 65, 1999).

played an increasingly important role in shaping public policy. The causal primacy of these factors in shaping the dimensions of the welfare state, however, appears to be greatly exaggerated.

A concomitant aspect of this exaggeration of global factors is the neglect of domestic forces of change—forces that are driven by technological advance and shifting demand patterns. These forces have caused massive shifts in the employment structure, the most notable being the shift from manufacturing to services. Because people often have skills that travel poorly between these sectors and because employer-provided social insurance is limited by a firm or by industry, deindustrialization poses significant risks that can be addressed only through government expansion of social security and public employment. In future work, the transferability of skills will be an important focus because the political effects of changes in the occupational structure depends on it. For example, a country like Germany with a training system that emphasizes specific skills will be politically more sensitive to occupational shifts than a country like the U.S. where the educational system emphasizes general skills.

Why has the role of deindustrialization been ignored in explanations of welfare state expansion? We suspect that one reason is a misconception that the shift in the employment structures is relatively uniform across countries, a common mistake in political science.[77] As we document in the introduction, there is in fact tremendous variation in the extent of deindustrialization, and our empirical results demonstrate that this factor can account for a very significant proportion of the variance in welfare state spending. Another reason for the omission is undoubtedly an outgrowth of the idea, deeply ingrained in most of our theories of comparative political economy, that the rise of the welfare state is linked to the strength of the industrial working class. What our analysis suggests is that any major transformation in the employment structure, whether from agriculture to industry or from industry to services, produces insecurities in the labor market that propel demands for state intervention.

Governments of all political stripes have responded to these demands by expanding transfer payments and social service provision. Nevertheless, partisanship continues to be important in the redistributive aspects of the welfare state. This shows up clearly in the results for public consumption, which has expanded much more rapidly in coun-

[77] Jonas Pontusson, "Explaining the Decline of European Social Democracy: The Role of Structural Economic Change," *World Politics* 47 (July 1995).

tries where the left is strong. We would conjecture that the same is true for transfer payments if the composition, rather than the level, of spending is examined.

In fact, there are reasons to expect that deindustrialization will be associated with increasingly distinct partisan effects. First, due to gaps in productivity growth across sectors, egalitarian policies tend to inhibit the expansion of private service sector employment. These policies present the government with an increasingly clear choice between either excluding more and more people from the labor market or employing more of them in public service sector jobs. Second, with the process of sectoral transformation coming to an end in many countries, the political support for further welfare state expansion is likely to wane, whereas distributive conflicts over existing welfare state programs are likely to intensify. We believe that these political aspects of deindustrialization are promising areas for future research.

Appendix: Glossary of Variables and Data Sources

$$\textit{automatic consumption} = \frac{\textit{gov consumption}}{GDP}(t-1) \bullet \left(\frac{\Delta \textit{gov deflator}(t)}{\textit{gov deflator}\,(t-1)}\right) / \left(\frac{\Delta GDP\ \textit{deflator}(t)}{GDP\ \textit{deflator}(t-1)}\right),$$

where gov deflator is the price deflator for government services, and GDP deflator is the price deflator for the whole GDP.

$$\textit{automatic transfers} = \textit{replacement rate}(t-1) \bullet \Delta \frac{\textit{unemployed} + \textit{population} > 65}{\textit{population}}(t),$$

where the replacement rate is the percentage share of transfers in GDP relative to the percentage share of the dependent population in the total population at time t − 1.[78]

capital formation. Gross capital formation as a percent of GDP.[79]

capital market openness. The index measures the extent to which capital markets are liberalized.[80]

[78] OECD, *Labour Force Statistics* (Paris: OECD, various years); and transfer data.

[79] Summers and Heston (fn. 74).

[80] Denis P. Quinn, and Carla Inclan, "The Origin of Financial Openness: A Study of Current and Capital Account Internationalization," *American Journal of Political Science* 41, no. 3 (1997).

deindustrialization. 100 minus the sum of manufacturing and agricul tural employment as a percentage of the working-age population.[81]

electoral participation. Based on voter turnout rates as recorded on ar annual basis in Mackie and Rose,[82] the *European Journal of Politica Research*, and in the International Institute for Democracy and Electoral Assistance.[83]

excluded labor. 100 minus all employed people as a percent of the total population (change in this variable when used as a dependent variable).

generosity of transfers. The percentage share of transfers in GDP relative to the percentage share of the nonworking population in the total population (change in this variable when used as dependent variable).[84]

government consumption. Total government consumption of goods and services net of military spending as a percentage of GDP.[85]

government transfers. All government payments to the civilian household sector, including social security transfers, government grants, public employee pensions, and transfers to nonprofit institutions serving the household sector.[86]

income. Gross domestic product per capita in purchasing power equivalents in 1985 U.S. dollars.[87]

left government center of gravity. This is an index of the partisan left-right "center of gravity" developed by Cusack.[88] It is based on (1) Castles and Mair's codings of government parties placement on a left-right scale,[89] weighted by (2) their decimal share of cabinet portfolios. The index varies from 0 (extreme right) to 4 (extreme left), although most observations are much closer to the mean.

private sector labor surplus. Deindustrialization minus the private service sector employment as a percent of the working-age population (change in this variable when used as dependent variable).

[81] OECD (fn. 78).

[82] Thomas T. Mackie and Richard Rose, *The International Almanac of Electoral History*, 3d ed. (London: Macmillan, 1991).

[83] International Institute for Democracy and Electoral Assistance, *Voter Turnout from 1945 to 1997: A Global Report on Political Participation* (Stockholm: IDEA Information Services, 1997).

[84] OECD (fn. 78); and transfer data.

[85] Thomas Cusack, "The Changing Contours of Government" (Wissenschaftszentrum Berlin für Sozialforschung, WZB Discussion Paper, no. 304, 1991); OECD (fn. 78); Stockholm International Peace Research Institute, *The SIPRI Year Book* (Stockholm: SIPRI, various years).

[86] Cusack (fn. 85); and OECD, *National Accounts, Part II: Detailed Tables* (Paris: OECD, various years).

[87] Summers and Heston (fn. 74).

[88] Cusack (fn. 24, 1997).

[89] Francis Castles and Peter Mair, "Left-Right Political Scales: Some 'Expert' Judgments," *European Journal of Political Research* 12, no. 1 (1984).

productivity growth. Annual rate of change in real value added per worker in industry and agriculture.[90]

strength of labor. Measured as the product of union density and centralization.[91]

trade balances. Merchandise trade balance expressed as a percent of GDP for three country groupings (OECD, OPEC, LDCs).[92]

trade openness. Total exports and imports of goods and services as percentage of GDP.[93]

unexpected growth. Real GDP per capita growth at time t minus average real per capita growth in the preceding three years. The variable is defined in accordance with Roubini and Sachs.[94]

[90] OECD, *National Accounts CD-Rom* (Paris: OECD, 1995)

[91] Density data: Jelle Visser, *European Trade Union in Figures* (Deventer, Netherlands: Kluwer Law and Taxation Publishers, 1989); and idem, "Unionization Trends Revisited" (Manuscript, University of Amsterdam, 1996). Centralization data: Torben Iversen, "Wage Bargaining, Central Bank Independence and the Real Effects of Money," *International Organization* 52 (Summer 1998).

[92] IMF, *Direction of Trade Statistics Yearbook* (Washington, D.C.: IMF, various years).

[93] OECD (fn. 86).

[94] Roubini and Sachs (fn. 23); and OECD (fn. 86).

WAGE INEQUALITY AND VARIETIES OF CAPITALISM

By DAVID RUEDA and JONAS PONTUSSON*

THE immediate goal of this article is to explore the determinants of wage inequality in advanced capitalist economies. Following a period in which the distribution of wages tended to become more compressed, most OECD countries have experienced some increase in wage inequality since 1980. However, the magnitude of change varies significantly across these countries: the rise of wage inequality began earlier and/or lasted longer in some countries than in others. As we shall see, the United States stands out as the OECD country that has experienced the most sustained rise of wage inequality, lasting at least a quarter of a century. With countries entering the 1980s at very different levels of wage inequality, the persistence of cross-national diversity remains a conspicuous feature of the data that we present below. In 1995 someone occupying the 90th percentile of the U.S. earnings distribution (the bottom of the top 10 percent) had an income that was 4.6 times larger than the income of someone in the 10th percentile (the top of the bottom 10 percent). At the opposite end of the spectrum, the 90–10 earnings ratio in Sweden was only 2.2 in 1995.

To date, most of the literature on the comparative political economy of labor markets has taken macroeconomic performance as the dependent variable and focused on the issue of wage restraint, or the trade-off between inflation and unemployment. In the corporatist tradition inspired by rational choice thinking, wage restraint is viewed as a public good, subject to familiar collective action problems, and divergent outcomes are typically explained in terms of institutional arrangements, which determine the ability of unions and/or employers to coordinate their wage-bargaining behavior. As we turn to explore wage-distributive outcomes, this line of thinking seems less compelling, for any number of (particular) wage distributions satisfy the conditions of Pareto

* This article has a long history: earlier versions have been presented in numerous forums, and a great many people have commented on the research presented here. We are especially grateful to Rob Franzese, Geoffrey Garrett, Torben Iversen, Walter Mebane, Michael Wallerstein, Chris Way, and Bruce Western for their constructive criticisms, technical assistance, and encouragement.

optimality. If the politics of wage restraint is essentially about coordination, trust, and perhaps sanctions to avoid a suboptimal outcome, the politics of wage distribution is more accurately described in terms of a continuous process of negotiating temporary settlements among competing interests.

In the real world the politics of wage restraint and the politics of wage distribution are, of course, opposite sides of the same coin, but as our analytical focus shifts from the former to the latter, we might expect that variables which capture the power resources of organized interests, such as union density and government partisanship, take on greater explanatory significance relative to more formal institutional variables, such as the degree of centralization of wage bargaining. Equally important, this shift of focus invites us to think differently about institutional arrangements. If centralization of bargaining matters to wage-distributive outcomes, this is not because it provides for coordination but rather because it affects the distribution of power among actors in the bargaining process.[1]

Quantitatively inclined comparativists are turning increasingly to pooled cross-section time-series regression analysis because it allows them to increase the total number of observations and to test relatively complex causal models with aggregate data from a small number of countries. Nevertheless, for all the sophistication of this recent quantitative work, more qualitatively inclined scholars still cling to the essential historical-institutionalist objection to regression analysis—that it presupposes but does not prove that the relationship between independent variable X and dependent variable Y is the same across all units of observation. The radical version of this objection holds that each country should be conceived as a unique context, determining the relationship between X and Y. More commonly in comparative political economy, we are told that countries cluster into a few broad, historically constituted institutional configurations. The arguments advanced by Katzenstein, Hall, Esping-Andersen, and Soskice imply that changing the value of X will have certain effects in one set of countries and quite different effects in another.[2]

[1] Cf. Michael Wallerstein, "Wage-Setting Institutions and Pay Inequality in Advanced Industrial Societies," *American Journal of Political Science* 43 (July 1999).

[2] Peter Katzenstein, *Small States in World Markets* (Ithaca, N.Y.: Cornell University Press, 1985); Peter Hall, *Governing the Economy* (New York: Oxford University Press, 1986); Gøsta Esping-Andersen, *The Three Worlds of Welfare Capitalism* (Princeton: Princeton University Press, 1990); David Soskice, "Wage Determination," *Oxford Review of Economic Policy* 6 (Winter 1990); and idem, "Divergent Production Regimes," in Herbert Kitschelt et al., eds., *Continuity and Change in Contemporary Capitalism* (New York: Cambridge University Press, 1999).

Behind our immediate goal of exploring the determinants of wage inequality, the ulterior purpose of this article is to show that regression analysis can and should incorporate the insights of the approach associated with these scholars.[3] Exploring the common determinants of wage inequality across nations and over time we seek to determine whether and how these determinants vary across varieties of capitalism. Drawing on a new data set collected by the OECD, we engage in two rounds of pooled regression analysis. In the first round we regress levels of wage inequality on unemployment rates, trade with low-wage countries, female labor-force participation, union density, centralization of wage bargaining, the public sector's share of total employment, and government partisanship. Although the first three variables are meant to capture supply-and-demand conditions, none of them turns out to be a consistent predictor of the observed variance in wage inequality; the other four variables, however, all have statistically and substantively significant coefficients. Our second regression setup adds further institutional complexity by distinguishing between social market economies (SMEs) and liberal market economies (LMEs).

To anticipate, our empirical results indicate that varieties of capitalism matter. We find some support for the proposition that SME conditions mute the impact of market forces on the distribution of wages. More importantly, our results show that SME conditions significantly affect either the direction or the magnitude of most of our political and institutional variables. Of particular interest to political scientists is the finding that the wage-distributive effects of government partisanship are contingent on institutional context. While leftist governments are associated with less wage inequality in liberal market economies, this is not the case in social market economies. Union density emerges as the single most important factor influencing wage inequality across institutional contexts; its effects are consistently egalitarian and they are greater than those of any other independent variable within the country clusters. Our results thus support the contention that the politics of wage distribution involves conflicts between unions and employers as well as distributive conflicts between different firms and different categories of wage earners.

[3] In seeking to incorporate institutional insights into regression analysis, we follow the lead of Bruce Western, *Between Class and Market* (Princeton: Princeton University Press, 1997); Geoffrey Garrett, *Partisan Politics in the Global Economy* (New York: Cambridge University Press, 1998); and Torben Iversen, *Contested Economic Institutions* (New York: Cambridge University Press, 1999). However, none of these works explicitly tests the idea that causal dynamics vary by political economy type. Previous work in this particular vein includes Philip O'Connel, "National Variations in the Fortunes of Labor," in Thomas Janoski and Alexander Hicks, eds., *The Comparative Political Economy of the Welfare State* (New York: Cambridge University Press, 1994); and Duane Swank, "Political Institutions and Welfare State Restructuring," in Paul Pierson, ed., *The New Politics of the Welfare State* (New York: Oxford University Press, 2000).

We begin by describing the wage-distributive outcomes that we seek to explain. We then review the literature on the determinants of wage inequality and generate causal hypotheses pertaining to the discrete independent variables identified above. Third, we elaborate on the distinction between social and liberal market economies and specify how we expect this distinction to affect our causal hypotheses. The fourth section briefly addresses methodological issues, and the fifth section presents our empirical results. By way of conclusion, we address the general implications of our analysis.

I. Patterns of Wage Inequality

The dependent variable in our analysis is a summary measure of the distribution of gross income from employment. The particular measure we use, the ratio of earning at the 90th percentile to earnings at the 10th percentile, is dictated by the OECD data set on which we rely. The 90–10 ratio, a measure of the distance between two points, certainly does not tell us everything about the overall shape of the distribution, but it is a commonly used measure and easy to interpret.

The reader should keep in mind that our inequality measure ignores important sources of income, such as self-employment, income from capital, and government transfers. It also ignores the distributive effects of taxation and income pooling within households. Moreover, the OECD data set on which we rely is restricted to full-time employees (except in the case of Austria). What follows, then, must not be confused with an analysis of the overall distribution of income in OECD countries. This said, income from employment accounts for the lion's share of income in all OECD countries, and the distribution of income from employment, as measured by 90–10 ratios, correlates quite closely with broader cross-national measures of income distribution.[4]

Figures 1 and 2 provide a graphic summary of the wage-distributive outcomes that our analysis seeks to explain. The figures reveal both persistent variations in levels of wage inequality among the sixteen OECD countries and considerable change over time. Starting with the persistence of cross-national variations, the U.S. and Canada clearly constitute a group unto themselves, distinguished by very high levels of wage inequality throughout the time period covered by our data. At the other

[4] Cf. Wallerstein (fn. 1), 649; OECD, *Income Distribution in OECD Countries: Evidence from the Luxembourg Income Study* (OECD, 1995); and Peter Gottschalk and Timothy Smeeding, "Cross-National Comparisons of Earnings and Income Inequality," *Journal of Economic Literature* 35 (June 1997). For further specifications of the wage data used here, see OECD, "Earnings Inequality, Low-Paid Employment and Earnings Mobility," *Employment Outlook* (July 1996).
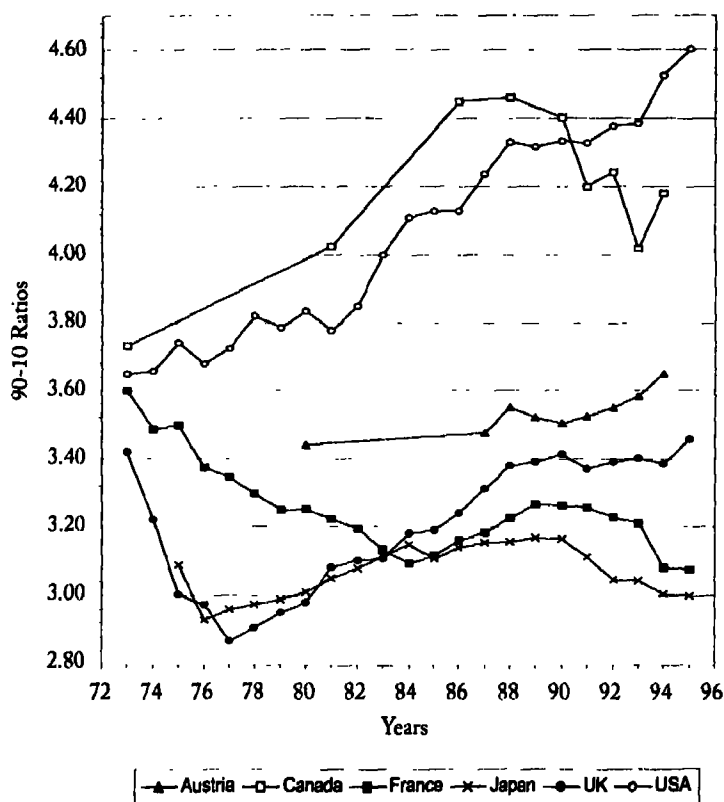
FIGURE 1

SOURCE: See appendix.

end of the spectrum the Scandinavian countries (Sweden, Norway, and Denmark) stand out as the OECD countries with the most egalitarian distribution of wages. The Scandinavian countries might be viewed as part of a broader low-inequality band that would also include Belgium, Finland, Italy, Switzerland, the Netherlands, Germany, and possibly Australia. Alternatively, the latter group of countries might be viewed as a separate, low-to-average band. In any case, Japan, France, the United Kingdom, and Austria make up another band, characterized by comparatively high levels of wage inequality, though not nearly as high as those of Canada and the U.S.[5] While relative rankings within these

[5] The high level of wage inequality in Austria is partly attributable to the fact that the underlying wage data include part-time employees, but other data sources also indicate that the Austrian distribution of wages is quite inegalitarian by continental European standards. See Bob Rowthorn, "Corporatism and Labour Market Performance," in Jukka Pekkarinen, Matti Pohjola, and Bob Rowthorn
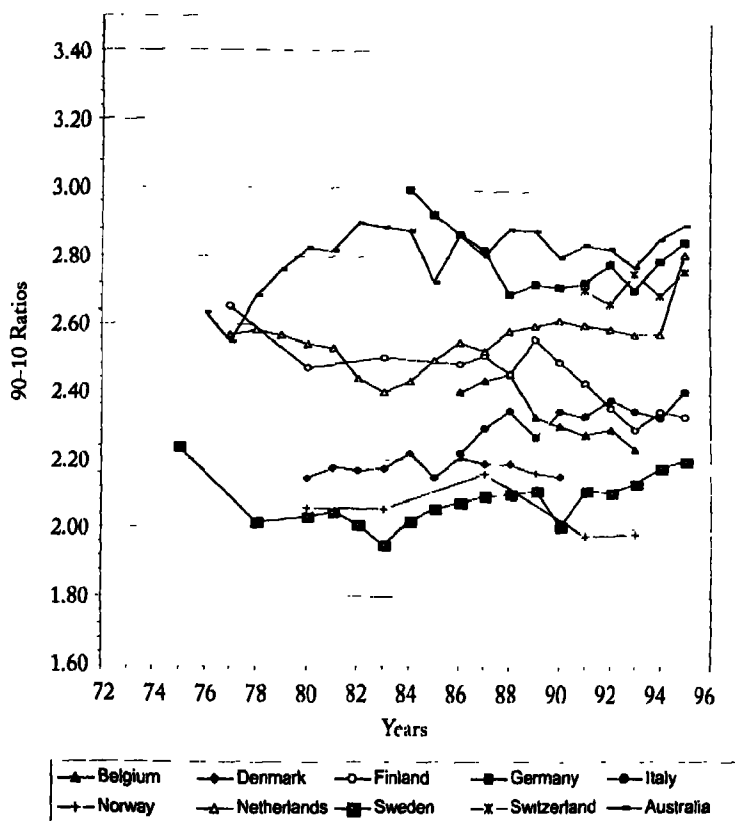
FIGURE 2

SOURCE: See appendix.

bands changed, the country composition of the bands remained remarkably stable over the 1973–95 period.

Looking at change over time, seven countries have experienced a continuous rise in wage inequality in recent years—most dramatically, the U.S. and the U.K, but also Austria, Germany, the Netherlands, Italy, and Sweden. In another four countries (Canada, France, Japan, and Australia), we observe notable increases of wage inequality in the 1970s and the 1980s, followed by subsequent decreases or, in the case of

eds., *Social Corporatism* (Oxford: Clarendon Press, 1992); and Jonas Pontusson, "Wage Distribution and Labor Market Institutions," in Torben Iversen, Jonas Pontusson, and David Soskice, eds., *Unions, Employers and Central Banks* (New York: Cambridge University Press, 2000). The use of country dummies and the small number of Austrian observations (7) renders our analysis essentially unaffected by the exceptional nature of the Austrian wage data.

TABLE 1
PERCENTAGE INCREASE IN 90–10 RATIOS FROM ALL-TIME TROUGH TO MOST RECENT OBSERVATION

Country	*% Change*	*Trough*	*Most Recent*
United States	26.0	1973	1995
United Kingdom	20.6	1977	1995
Netherlands	17.5	1983	1995
Australia	14.1	1977	1995
Sweden	12.8	1983	1995
Canada	12.1	1973	1994
Italy	8.6	1986	1995
Germany	6.3	1988	1995
Austria	6.1	1980	1994
Switzerland	3.7	1992	1995
Japan	2.4	1976	1995
Finland	1.7	1993	1995
Belgium	0.0	1993	1993
Denmark	0.0	1990	1990
France	0.0	1995	1995
Norway	0.0	1993	1993

SOURCE: See appendix.

Australia, a period of stability. Norway, Denmark, and Switzerlan seem best characterized as cases of long-term stability while Finlan and Belgium stand out as the two countries in which wage inequalit declined continuously in the 1980s and 1990s. Measuring the percent age change in 90–10 ratios from their all-time trough to the most re cent observation available, Table 1 brings out the widespread tren toward increased wage inequality, as well as important cross-nationa variations in its duration and magnitude.[6]

II. LITERATURE REVIEW AND HYPOTHESES

The labor economics approach to the problem of explaining wage distributive outcomes focuses on relative demand for and supply of dif

[6] The trend for wage inequality becomes more impressive if we look at the wage distribution amon men and women separately. See OECD (fn. 4, 1996); and Jonas Pontusson, David Rueda, and Christo pher Way, "The Role of Political-Institutional Variables in the Making of Gendered Patterns of Wag Inequality" (Working paper, Institute for European Studies, Cornell University, 1999). In most coun tries increases of within-gender inequality have been offset by continued compression of between gender wage differentials.

ferent types of labor. Looking at the demand side, labor economists typically argue that the growth of wage inequality in the U.S. and elsewhere reflects technological changes, which have rendered more-educated workers more valuable to employers than less-educated workers. With regard to the supply side, labor economists frequently point out that the compression of wage differentials prior to the 1970s coincided with rapid growth of university enrollments, that is, with an increase in the relative supply of educated labor. As the growth of university enrollments decelerated, the supply of better-educated labor subsequently failed to keep up with demand, giving rise to sharply increasing returns to education. At the same time, immigration and, more broadly relevant to all OECD countries, the massive increase in women's participation in the labor force since the 1970s represent an increase in the relative supply of unskilled labor to the extent that immigrants and women are on average less educated and less experienced than "native" men.[7]

In a similar vein, Wood argues that much of the trend toward increased wage inequality in the OECD countries in the 1980s can be attributed to increased manufacturing trade with less developed countries.[8] Though his empirical analysis is controversial, the logic of his argumentation is consistent with the supply-and-demand framework of the labor economics approach: as imports of less skill-intensive goods from low-wage countries increase, the effective supply of unskilled labor relative to skilled labor increases, putting downward pressure on the relative wages of unskilled workers.

Labor economists who engage in cross-national comparison typically find that supply-and-demand factors alone cannot explain observed variations in wage inequality across the OECD countries; they conclude that institutions matter.[9] When economists speak of "institutions," they have in mind not only codified rules or formalized organizational arrangements but also government policy and the distribution of power among organized interests. Commonly, the sectoral distribution of employment and other dimensions of industrial structure are also referred to as institutional variables. From the perspective of comparative political economy, the interesting question is not whether in-

[7] Robert Topel, "Wage Inequality and Regional Labor Market Performance in the United States," in Toshiaki Tachibanaki, ed., *Labour Market and Economic Performance* (New York: St. Martin's Press, 1994); and Lars Svensson, *Closing the Gender Gap* (Lund: Ekonomisk-Historiska Föreningen, 1995).

[8] Adrian Wood, *North-South Trade, Employment and Inequality* (Oxford: Clarendon Press, 1994).

[9] Gottschalk and Smeeding (fn. 4); Rowthorn (fn. 5); Francine Blau and Lawrence Kahn, "International Difference in Male Wage Inequality," *Journal of Political Economy* 104 (August 1996); and various contributions to Richard Freeman and Lawrence Katz, eds., *Differences and Changes in Wage Structures* (Chicago: University of Chicago Press, 1995).

stitutions in this broad sense matter but rather which institutions matter and how (why) they matter. In this spirit, we seek to isolate the causal effects of different political-institutional variables and to explore interactions among those variables.

As a first cut, we derive four relevant political-institutional variables from the literature: unionization, centralization of wage bargaining, the size of the public sector, and the partisan composition of government. Ideally, we would want to control for all the relative supply-and-demand factors identified by labor economists as we estimate the effects of these variables. Measuring relative demand shifts associated with new technology is notoriously difficult, however, and we have not been able to find any such measure that would suit our purposes. Moreover, the existing OECD data on immigration flows are not comparable across countries and OECD data on educational qualifications exist only for the second half of the time period covered by our wage-inequality data.

Thus we are left with two control variables meant to capture cross-national variations and variations over time in the relative supplies of skilled and unskilled labor: non-OPEC LDC trade as a percentage of GDP and women's share of total employment. For the reasons set out above, we hypothesize that both variables are positively associated with wage inequality. With respect to female labor-force participation, some preliminary qualifications are in order, for we know the 1970s and early 1980s were characterized by a significant increase of female labor-force participation, as well as a significant decline of wage inequality in many OECD countries. Our baseline hypothesis implies that other factors, such as increased female participation in higher education and the expansion of public sector jobs, offset the inegalitarian effects of female labor-force participation in this period, but it is also plausible that the wage-distributive effects of female labor-force participation itself are contingent or contradictory. As women acquire skills through labor-force participation, higher levels of female labor-force participation should be associated with a smaller skill gap between men and women.

The regressions reported below also include the rate of unemployment to capture the wage-distributive effects of aggregate demand fluctuations (as distinct from relative demand shifts). It is a commonplace that unskilled, low-paid workers are more readily substitutable than skilled, high-paid workers and that their bargaining position is more immediately and more adversely affected by unemployment. By this logic, we expect the rate of unemployment to be positively associated with wage inequality.[10]

[10] Cf. James Galbraith, *Created Unequal* (New York: Free Press, 1998).

However, there is another side to the relationship between unemployment and wage inequality, for employers are more likely to lay off unskilled workers than to lay off skilled workers during economic downturns. To the extent that an increase in unemployment entails a disproportionate loss of low-paid jobs, it should be associated with less rather than more wage inequality (though it would still be associated with more overall inequality of income). For both female labor-force participation and unemployment, then, our theoretical expectations are somewhat ambiguous. Still, it is desirable to include these variables as controls. Let us now turn to the variables of primary theoretical interest.

UNION DENSITY

Following Freeman, we should distinguish two dimensions of the relationship between unionization and wage distribution: one concerns the distribution of wages among union members and how it compares with the distribution of earnings among unorganized wage earners; the other concerns wage differentials between union members and nonmembers, that is, the wage premium associated with union membership for wage earners with equivalent qualifications, experience, and other relevant characteristics.[11] With respect to the first dimension, there are several reasons to expect the wage distribution of the union sector of an economy to be more compressed than that of the nonunion sector. To begin with, unions approximate the logic of democratic decision making (one person, one vote) more closely than markets do, and whenever the mean wage exceeds the median wage, we would expect a majority of union members to favor redistributive wage demands. Moreover, because unions are dependent on membership support in conflicts with management, they have a strong interest in curtailing wage setting based on the subjective decisions of foremen or personnel managers.

While employers competing in the same product markets can be expected to favor standardization of wage rates across firms, thereby taking wages out of competition, they also have an interest in maintaining unilateral control of intrafirm differentials.[12] Moreover, union interest in interfirm standardization is probably deeper and more general than that of employers. Even when market conditions differ among firms, so that price discrimination by the union is possible, the need to sustain mobilizational capacity inclines unions to restrict the scope of interfirm differentials.

[11] Richard Freeman, "Unionism and the Dispersion of Wages," *Industrial and Labor Relations Review* 34 (October 1980); idem, "Union Wage Practices and Wage Dispersion within Establishments," *Industrial and Labor Relations Review* 36 (October 1982); cf. also Blau and Kahn (fn. 9).

[12] Cf. Freeman (fn. 11, 1980).

To the extent that unionized workers earn more than equivalent nonunionized workers, the relationship between unionization and wage distribution becomes more complicated, for the wage-distributive effects of unionization now come to depend in part on the distribution of union membership across the wage hierarchy. Unionism would be a source of wage inequality if highly paid wage earners were better organized than low-paid workers, and the opposite would hold if low-paid wage earners were better organized. Though we lack detailed information on the distribution of union membership by income, the available evidence suggests that in most countries any wage premiums that accrue to union members are likely to compress the wage distribution.[13] Overall, then, we expect union density to be negatively associated with wage inequality.

Wage-Bargaining Centralization

A number of recent studies establish that countries with more centralized wage-bargaining systems consistently tend to have a more compressed distribution of wages than countries with less centralized wage-bargaining systems.[14] Centralization facilitates the reduction of interfirm and intersectoral wage differentials, since it means that more firms and sectors are included in a single wage settlement, but this argument presupposes that at least one of the parties of centralized bargaining wants to achieve a reduction of interfirm or intersectoral differentials. Wallerstein articulates two mechanisms whereby centralization produces egalitarian outcomes: there is a political mechanism through which centralization alters "the influence of different groups in the wage-setting process," and then there is an ideological mechanism whereby centralization affects norms of fairness.[15]

In the paradigmatic Swedish case, low-wage affiliates of the powerful confederation of blue-collar unions (Landsorganisationen) insisted on solidaristic measures as a condition for their participation in peak-level bargaining sought by employers in the 1950s.[16] But why should centralization systematically strengthen the relative bargaining power

[13] The Eurobarometer of June–July 1994 yields the following figures for average union density by income quartile in the EU member states: lowest quartile, 37.5 percent; second-to-lowest quartile, 37.8 percent; second-to-highest quartile, 34.2 percent; and highest quartile, 23.7 percent. For all countries but Canada the union-density data used in our regressions refer to net rather than gross union density ("employed union members as a percentage of the employed labor force" rather than "union members as a percentage of the total labor force").

[14] Wallerstein (fn. 1); Iversen (fn. 3), 2–5; Rowthorn (fn. 5); Blau and Kahn (fn. 9); and OECD, "Economic Performance and the Structure of Collective Bargaining," *Employment Outlook* (July 1997).

[15] Wallerstein (fn. 1), 674.

[16] Peter Swenson, *Fair Shares* (Ithaca, N.Y.: Cornell University Press, 1989), 56–58.

of low-wage unions? Wallerstein's political explanation essentially reproduces Freeman's argument about a single union that formulates wage demands on the basis of some form of majoritarian decision making: if low-wage and high-wage unions bargain jointly, organizational politics will influence the demands that they pursue, and market forces will be less influential in determining the distribution of wage increases. Consistent with Wallerstein's ideological explanation, we also hypothesize that centralized bargaining—in the extreme, a single settlement for all wage earners—renders wage differentials more transparent and thus politicizes wage-distributive outcomes. By this logic, centralization not only empowers low-wage unions but also makes them more likely to demand redistributive measures.

To test these hypotheses, which imply a negative association between centralization and wage inequality, we rely on an index of wage-bargaining centralization developed by Iversen.[17] This index takes into account the degree of union concentration at different levels of wage bargaining, as well as the relative importance of local, sectoral, and national bargaining.[18]

Government Employment

Kahn observes a negative association between the size of the public sector (relative to total employment) and wage inequality and explains this association by arguing that countries expand government employment in response to the employment-dampening effects of wage compression.[19] By contrast, we want to suggest that the size of the public sector may be construed as a cause of wage-distributive outcomes. Conceived in this fashion, the relationship between government employment and the overall distribution of wages is twofold, just like the relationship between unionization and wage distribution: first, there is the question of how the distribution of wages in the public sector compares with the distribution of wages in the private sector, and second there is the question of wage differentials between these sectors and how they affect the overall distribution of wages.

There are several reasons to expect wages in the public sector to be more compressed than wages in the private sector, aside from the fact

[17] For a detailed specification, see Iversen (fn. 3), 83–86.

[18] To capture the inertia associated with institutional change, these yearly figures were lagged so that the value for a given year used in our regressions is the average for that year and the previous four years. It should also be noted that we have extrapolated centralization values for the last two years of our time series.

[19] Lawrence Kahn, "Wage Inequality, Collective Bargaining and Relative Employment 1985–94" (Working paper, Institute for Labor Market Policies, Cornell University, 1999).

that the public sector is more heavily unionized than the private sector in most OECD countries.[20] On average, public sector unions appear to be more inclined than their private sector counterparts to favor wage solidarity. This generalization certainly holds for the 1970s and 1980s, when public sector employers also appear to be more inclined than private sector employers to accommodate union demands for compression or even to initiate compression. While sheltered from competition in product markets, public sector employers are more directly exposed to political pressures in an egalitarian direction, being directly accountable to elected officials. By the same token, public sector unions have had less reason to worry about any potential trade-off between wage compression and employment growth than their private sector counterparts have had.[21]

As with union density, then, we expect the size of the public sector to be negatively associated with wage inequality because wage differentials are more compressed within the public sector. However, public sector wage premiums are much less likely to be egalitarian than are union wage premiums.[22] After all, government employees include a great many well-educated and highly paid civil servants and other professionals. Hence our expectations for the overall impact of the size of the public sector on wage inequality are somewhat uncertain. It seems likely that the effects of this variable are contingent on other institutional variables.

Government Partisanship

There are good reasons to expect leftist parties to pursue redistributive policies when they hold government power.[23] The effects of government partisanship will manifest themselves primarily in terms of redistribution via government taxation and spending, but government policies also affect the distribution of market incomes in general and of wages in particular. Indirectly, government policies affect the distribution of wages via their effects on unemployment, the size of the public sector, and union density. More directly, governments influence the distribution

[20] OECD, "Trends in Trade-Union Membership," *Employment Outlook* (July 1991), 113.

[21] Focusing on wage restraint rather than wage distribution, Geoffrey Garrett and Christopher Way present similar arguments about the distinctive dynamics of public sector bargaining; see Garrett and Way, "Public Sector Unions, Corporatism and Wage Determination," in Iversen, Pontusson, and Soskice (fn. 5).

[22] Drawing on microdata for the early 1990s from the Luxembourg Income Survey, Gornick and Jacobs report very sizable public sector wage premiums for men and women alike in Belgium, Canada, Germany, the Netherlands, the U.K., and the U.S., but not in Sweden. See Janet Gornick and Jerry Jacobs, *Gender, the Welfare State and Public Employment*, Luxembourg Income Study Working Paper no. 168 (1997).

[23] Cf., e.g., Garrett (fn. 3).

of wages through minimum-wage and equal-pay legislation, other forms of income policy, and a variety of measures that strengthen the competitive position of women and other disadvantaged groups (for example, immigrants) in the labor market. Partly for lack of good quantitative measures of such policies, we test the partisanship hypothesis by including Cusack's index of the cabinet center of gravity in our regressions.[24] As this index ranges between a score of 1 for a pure government of the radical left to a score of 5 for a pure government of the radical right, we expect the index values to be positively associated with wage inequality.

III. THE TWIST: VARIETIES OF CAPITALISM

As indicated at the outset, this article explores whether the idea that the advanced capitalist countries form clusters with distinctive causal dynamics is applicable to the problem of explaining wage inequality. The typology of advanced capitalist countries used here hinges on the distinction between social market economies (SMEs) and liberal market economies (LMEs). The Northern European countries referred to as "social market economies" are essentially the same as the countries Katzenstein refers to as "corporatist," but the substantive connotations of the SME concept are different from those of the corporatism concept.[25] Katzenstein's well-known definition of corporatism emphasizes the formation of government policy through bargaining among centralized interest groups and the ideology of social partnership. Following Soskice, our conceptualization focuses more directly on the regulation of markets (government outputs rather than inputs).[26] However, our distinction between social and liberal market economies is narrower than Soskice's distinction between coordinated and liberal market economies and therefore yields a somewhat different grouping of countries.

Three basic features distinguish social market economies from liberal market economies. First, social market economies are characterized by comprehensive, publicly funded social welfare systems. Though the degree of redistribution varies, public welfare programs provide a relatively high "reservation wage" for the jobless and reduce workers' dependence on particular employers by providing for portability of employment benefits, retraining opportunities, sick pay insurance, and parental leave insurance. Using Esping-Andersen's terminology, we

[24] Tom Cusack, "Partisan Politics and Public Finance," *Public Choice* 91, no. 3–4 (1997).

[25] Katzenstein (fn. 2); and idem, *Policy and Politics in West Germany* (Philadelphia: Temple University Press, 1987).

[26] E.g., Soskice (fn. 2, 1990 and 1999).

might say that all these countries have reached a critical threshold · labor decommodification via the public provision of social welfare.[27]

Second, social market economies are characterized by governme regulation to standardize employment conditions and to provide for high degree of employment security. The details of such labor-mark regulation vary from one country to another, but the effects are broad similar: increased costs for employers to shed labor and greater stai dardization of employment conditions across sectors and categories · labor. While pay scales may be more or less compressed, overt pay di crimination based on gender, race, or legal status (in the case of imm grants) is less common in social market economies than in liber market economies. The employment conditions of full-time and par time employees are also more alike.

Third, social market economies are distinguished by a high degree · institutionalization of collective bargaining and coordination of wa formation. By coordination of wage formation, we mean that wage d velopments in different sectors of the economy are more tightly co pled than they are in liberal market economies. As suggested by tl Japanese case, employers may be able to coordinate the wage-formatic process by themselves under certain circumstances: what is distincti about the social market economies is that coordination occurs throug collective bargaining, which gives unions a central role in the process

While the public provision of welfare and government regulation · employment conditions clearly introduce new considerations into o analysis, the institutionalization of collective bargaining may appear be nothing but the sum of union density and wage-bargaining centra ization. These variables are undoubtedly correlated on a cross-nation basis, but they are conceptually distinct. Union density pertains to tl balance of power between unions and employers, centralization pe tains to the formal organization of the wage-bargaining process, ar institutionalization describes the influence of collectively bargaine wages on actual wages across the whole economy. Estimated by rece OECD studies, the percentage of the labor force covered by collecti bargaining agreements provides a rough and ready measure of instit tionalization. Controlling for union density, coverage rates tell us abo the extent to which nonunion workers are affected by the employme terms achieved by unions through collective bargaining.[28]

[27] Esping-Andersen (fn. 2).

[28] Collective bargaining coverage rates exceed union density either because governments decide extend negotiated agreements to firms or sectors that were not party to the agreements (the Fren case) or because employers are better organized than unions and collective agreements encompass employees of the firms that are party to the agreement (the German case).

TABLE 2
FEATURES DISTINGUISHING SOCIAL AND LIBERAL MARKET ECONOMIES
(INDEX SCORES OR PERCENTAGES WITH OVERALL RANKS IN PARENTHESES)

	Welfare State		*Collective-Bargaining Coverage 1990*	*Employment Protection 1989*
Country	*Decom 1980*	*Spend/GDP 1990*		
SMEs				
Austria	31.1 (6)	23.9 (8)	98 (1)	9.0 (7)
Belgium	32.4 (5)	26.6 (5)	90 (4)	10.5 (3)
Denmark	38.1 (3)	28.3 (3)	69 (11)	3.3 (11)
Finland	29.2 (8)	25.3 (7)	95 (2)	10.5 (3)
Germany	27.7 (9)	23.8 (9)	90 (4)	12.0 (2)
Netherlands	32.4 (4)	28.8 (2)	71 (10)	7.3 (9)
Norway	38.3 (2)	26.9 (4)	75 (9)	9.8 (5)
Sweden	39.1 (1)	32.6 (1)	83 (6)	8.5 (8)
SME average	33.5	27.0	83.9	8.9
LMEs				
Australia	13.0 (16)	13.1 (15)	80 (8)	3.3 (11)
Canada	22.0 (14)	18.0 (12)	38 (14)	1.7 (15)
Japan	27.1 (11)	12.4 (16)	23 (15)	3.7 (10)
Switzerland	29.8 (7)	17.4 (13)	50 (12)	1.8 (14)
UK	23.4 (13)	19.8 (11)	47 (13)	2.3 (13)
USA	13.8 (15)	14.2 (14)	18 (16)	0.4 (16)
LME average	21.5	15.8	42.7	2.2
MIXED				
France	27.5 (11)	26.0 (6)	92 (3)	9.5 (6)
Italy	24.1 (12)	23.0 (10)	82 (7)	14.3 (1)

SOURCES: Decommodification index: Esping-Andersen (fn. 2), 52; total social spending as a percentage of GDP: OECD, *New Directions in Social Policy in OECD Countries* (Paris: OECD, 1994); and idem, *Social Expenditure Statistics in OECD Member Countries*, Labour Market and Social Policy Occasional Papers, no. 17 (1996); collective bargaining coverage rates: OECD (fn. 14), 71; employment protection index: OECD, *The OECD Jobs Study: Evidence and Explanations*, pt. 2 (Paris: OECD, 1994), 74.

Table 2 provides quantitative indicators of the three features that distinguish SMEs from LMEs for our sixteen countries. These figures are meant to be illustrative: in our conceptualization the distinction between social and liberal market economies is a categorical one and cannot be derived simply from the sum of these measures. To be categorized as a social market economy, moreover, a country must partake of all three features of "SME-ness."

Drawing on the typologies of other scholars, as well as on the data presented in Table 2, the majority of our countries are easily categorized as either SMEs or LMEs. France and Italy are the exceptions. Thus, al-

though Italy's welfare state is smaller and less decommodifying than the countries we categorize as SMEs, it has the strictest employment protection legislation of all the countries included in our analysis, and its level of collective-bargaining coverage exceeds that of several SMEs. Ranking sixth in terms of social spending as a percentage of GDP, France would seem to have an even better claim to be coded as an SME. In view of the disorganized and fragmented character of French industrial relations, however, we do not believe that collective-bargaining coverage represents a meaningful measure of institutionalization in the French case.[29] While Italy lacks the welfare-state characteristics associated with SME-ness, France falls short when it comes to the institutionalization of collective bargaining. Nevertheless, it would surely be misleading to categorize these countries as liberal market economies. Following Hall, therefore, we resolve this problem by creating a separate category—"mixed economies"—for France and Italy.[30]

The typology presented in Table 2 diverges most clearly from conventional wisdom in its assignment of Japan and Switzerland to the category of liberal market economies. With respect to Japan, the reader must keep in mind that the criteria underlying our typology pertain specifically to government regulation of labor markets and/or joint regulation by employers and unions. Unquestionably, the Japanese case fails to conform to the model of a liberal market economy in matters of finance, corporate governance, and industrial policy. As for Switzerland, its score on Esping-Andersen's decommodification index falls well within the range of the social market economies, but on our other three indicators it clearly falls within the range of the liberal market economies.[31]

Table 3 summarizes how we expect the distinction between social and liberal market economies to affect the causal relationships hypothesized in the previous section. In general, we expect SME conditions to mute or neutralize the inegalitarian impact of market forces, measured here by unemployment, LDC trade, and female labor-force participation. The argument about unemployment is perhaps the most obvious one: if laid-off workers receive generous unemployment compensation and other welfare benefits, an increase in unemployment is less likely to

[29] The coverage rates estimated by the OECD do not necessarily pertain to wage agreements. A government decision to extend a collective agreement on summer holidays could account for France's exceptionally high coverage rate.

[30] Peter Hall, "Organized Market Economies and Unemployment in Europe" (Manuscript, Center for European Studies, Harvard University, 1998).

[31] As we explain in the conclusion, the question of how to code Japan and Switzerland turns out to be of no consequence for our main findings.

TABLE 3
SUMMARY OF EXPECTED ASSOCIATIONS WITH WAGE INEQUALITY, TAKING SME/LME INTERACTIONS INTO ACCOUNT

Variable	*SMEs*	*LMEs*
Unemployment	–/0	0/+
LDC trade	0/+	+
Female labor-force participation	0	+
Union density	–	–
Bargaining centralization	–	–/0
Government employment	–	0
Government partisanship		
hypothesis 1	0	+
hypothesis 2	–	0

exert downward pressure on wages at the lower end of the wage hierarchy. With respect to LDC trade, we proceed from the proposition that LMEs are more segmented than SMEs. In a segmented economy, an increase of LDC imports will put downward pressure on the wages of the workers who produce the goods that are exposed to new competition from low-wage producers. In a more closely integrated or coordinated economy, we expect low-wage competition to translate into cuts in real wages for all workers, rather than into increased wage inequality. Finally, SME conditions are likely to alter the wage-distributive consequences of female labor-force participation, because women tend to be less skilled than men and often end up in precarious forms of employment for other reasons as well (most obviously child-rearing responsibilities). Outside the public sector, feminization of the workforce typically correlates with low levels of unionization. As noted above, the institutional arrangements that distinguish SMEs from non-SMEs serve to standardize employment conditions and provide for the extension of union-negotiated wage contracts to nonunion members. This should mitigate the inegalitarian effects of an increase of female labor-force participation.

Turning to the political-institutional variables that constitute our primary interest, the institutional arrangements characteristic of SMEs enhance the significance of collective bargaining relative to market forces and, as a result, the characteristics of collective bargaining should matter more to wage-distributive outcomes. In SMEs, where reservation wages are high and bargaining coverage is extensive, the equalizing effect of bargaining centralization should be especially large.

The implications of SME conditions for the wage-distributive impact of unionization are less straightforward. On the one hand, one might suppose that those aspects of social market economies that enhance employment security would serve to desensitize unions and their members to the potential employment costs of wage compression and therefore render them more prone to pursue (support) egalitarian wage demands. By this logic, a one-unit increase in unionization should be associated with a larger reduction of wage inequality in SMEs than in LMEs. On the other hand, the extension of union-negotiated wage contracts to nonunion workers would seem to imply that levels of wage inequality in SMEs are less sensitive to changes in union density than they are in LMEs. As indicated in Table 3, our prior assumption is that these countervailing effects cancel each other out and that the wage-distributive effects of union density are not affected by the SME-LME distinction.

With respect to government employment, we expect that its egalitarian impact will be greatest under SME conditions. In both types of economies, private sector employers who compete with public sector employers for labor at the lower end of the labor market must follow suit if public sector employers raise the relative wages of the lowest paid employees. By linking wage developments in different sectors more closely, the institutional arrangements characteristic of SMEs should expand the spillover of public sector wage compression into the private sectors.

Finally, there are two competing ways to think about the interaction between social market conditions and government partisanship. On the one hand, SME arrangements might be conceived as constraints that limit the effects of partisanship (or ideology) on government policy. On the other hand, governments would seem to be significantly more involved in wage formation in SMEs than in non-SMEs, and therefore the preferences of governing parties may have more important consequences for wage formation. According to the latter perspective, the association between leftist government and wage compression should be stronger in SMEs than in non-SMEs; according to the former perspective, it should be weaker.

IV. Methodology

In pooled cross-section time-series analysis, "country-years" are the units of observation of dependent and independent variables. By incorporating over-time variations, pooling dramatically increases the total number of observations and enables us to test more complex causal

models against data from a relatively small number of countries. At the same time this methodology is inextricably linked to the idea that cross-national variations and changes over time have common determinants; more precisely, it is linked to the goal of ascertaining the common determinants of cross-national variations and changes over time.

We address the autocorrelation problem associated with time-series data by including the lagged dependent variable on the right-hand side of the equation. Following Beck and Katz, we include a lagged dependent variable because it "makes it easier . . . to examine dynamics and allows for natural generalizations in a manner that the serially correlated errors approach does not."[32] As the distribution of wages changes only marginally from one year to the next, the coefficient for levels of wage inequality in the previous year turns out to be highly significant. On the assumption that the effects of a one-unit change in a particular variable persist, the long-term effects of such a change can be computed by dividing the value of the coefficient for the variable of interest by 1 minus the coefficient for the lagged dependent variable.[33] In what follows, we estimate both long-term and immediate effects for each variable.

Our regression models also include dummy variables for each of the countries in our data set. We do this to eliminate the bias resulting from the effects of country-specific omitted variables.[34] Put somewhat crudely, the country dummies control for the values that all observations for a given country share. While we are not interested in why particular countries fit our regression more or less closely and therefore do not report the coefficient estimates for the country dummies below, we believe that it is essential to control for country-specific effects in this manner. Scholars engaged in cross-national comparison sometimes eschew the use of country dummies on the grounds that they simply tell us that countries are different, when the interesting question is how or why they are different. Yet there is every reason to suspect that outcomes such as ours are influenced by country-specific historical or cultural factors, which cannot be measured on a cross-national basis (for instance, the influence of milltown culture on the priority assigned to wage solidarity by the Swedish labor movement). The results of F-tests

[32] Nathaniel Beck and Jonathan Katz, "Nuisance vs. Substance," in John Freeman, ed., *Political Analysis* (Ann Arbor: University of Michigan Press, 1996), 6:1. The results of Breusch-Godfrey tests indicate that there is no significant autocorrelation in our regressions. In tests with a variety of lags, we could not reject the null hypothesis (nonexistence of autocorrelation) at a level even close to the 90 percent traditional significance threshold. See William Greene, *Econometric Analysis* (Englewood Cliffs, N.J.: Prentice Hall, 1997), 595.

[33] Cf. George Box and Gwilyn Jenkins, *Time Series Analysis* (Oakland, Calif.: Holden-Day, 1976), chap. 1.

[34] See Cheng Hsiao, *Analysis of Panel Data* (New York: Cambridge University Press, 1986).

strongly confirm that country dummies belong in the specification of our regression models.[35]

Nickell demonstrates that, with short panel data, OLS estimation of models with lagged dependent variables and fixed effects produces biased coefficients.[36] We address this problem by computing the instrumental variable (IV) estimator suggested by Anderson and Hsiao and using a two-stage IV procedure.[37]

In what follows, we present first the results of a linear regression model and then the results of a model in which our independent variables are interacted with dummy variables for social market economies (SME), liberal market economies (LME), and mixed economies (MIX). We use the independent variables, a one-year lag of the independent variables, and the country dummies as the instruments and treat the lagged dependent variable as endogenous.[38] For the linear model we estimate the following equation:

$$y_{it} = \chi \hat{y}_{i,t-1} + \sum_k \beta_k x_{kit} + \alpha_i + \varepsilon_{it} \qquad (1)$$

Where y_{it} is the dependent variable, the x's are the independent variables, i refers to the cross-sectional units, t to the time units, k to the number of independent variables, α_i refers to the separate intercepts for each country (there is no common intercept), β_k refers to the slopes of the explanatory variables, χ to the slope of the lagged dependent variable predicted in the first stage of the IV procedure, and ε_{it} is a random error term normally distributed around a mean of 0 with a variance of σ^2.

The setup of our interaction model is the same except that we introduce three additional dummy variables (SME, LME, and MIX), which we interact with the x variables. Again we use the independent variables, a one-year lag of the independent variables and the country dummies as the instruments and treat the lagged dependent variable as endogenous. We estimate the following equation:

[35] For the regressions reported in Tables 2 and 3 the results of F-tests show that the country dummies are significant at better than the 99 percent level.

[36] Stephen Nickell, "Biases in Dynamic Models with Fixed Effects," *Econometrica* 49 (November 1981).

[37] See T. W. Anderson and Cheng Hsiao, "Estimation of Dynamic Models with Error Components," *Journal of the American Statistical Association* 76 (September 1981); and idem, "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics* 18 (January 1982). This is similar to the method suggested by Hibbs and employed by Alvarez, Garrett, and Lange. See Douglas Hibbs, "Problems of Statistical Estimation and Causal Inference in Dynamic Time Series Models," in Herbert Costner, ed., *Sociological Methodology 1973/1974* (San Francisco: Jossey-Bass, 1974); and Michael Alvarez, Geoffrey Garrett, and Peter Lange, "Government Partisanship, Labor Organization and Macroeconomic Performance," *American Political Science Review* 85 (June 1991).

[38] In both the linear and the interaction models, the instruments we use turn out to be excellent predictors of the lagged dependent variable. The R^2 obtained in the first-stage IV regressions was higher than .95.

$$y_{it} = \gamma\hat{y}_{i,t-1} + \sum_k \lambda_k * x_{kit} * \text{SME} + \sum_k \omega_k * x_{kit} * \text{LME} + \sum_k \delta_k * x_{kit} * \text{MIX} + \tau_i + \eta_{it} \quad (2)$$

where τ_i refers to the separate intercepts for each country (there is no common intercept), λ_k, ω_k, and δ_k refer to the slopes of the explanatory variables, γ to the slope of the lagged dependent variable predicted in the first stage of the IV procedure, and η_{it} is a random error term normally distributed around a mean of 0 with a variance of σ^2.[39]

For each x variable, this equation yields separate coefficient estimates for the countries we have coded as social market economies, liberal market economies, and mixed cases. The same coefficient estimates would be obtained if the linear model were run separately for each set of countries, but the results of our interaction model are more directly comparable to the results of our linear model, since the total number of observations is the same in the two models.

Two other features of our analysis should be noted before we turn to the empirical results. First, the analysis includes interpolated observations of wage inequality for some of the years with missing observations in Figures 1 and 2. Observations for missing years were interpolated on the assumption that any change between two years was evenly distributed across the intervening years, but we decided not to interpolate data across more than three missing years. The analysis therefore uses only 20 interpolated observations out of a total of 217.[40] Second, we engaged in logarithmic transformations of all variables other than the dummy variables before running the regressions. When variables on both sides of the regression equation are logged, the regression coefficients can be interpreted as measuring the percentage change in the dependent variable associated with a percentage change in the independent variable. The slope values yielded by our regressions thus represent elasticity measures of the relationship between the variables.

[39] Nathaniel Beck and Jonathan Katz propose a method for deriving consistent standard error estimates in the presence of panel-heteroscedastic errors that has been widely adopted by students of comparative political economy, e.g., Garrett (fn. 3); and Iversen (fn. 3); see Beck and Katz, "What to Do (and Not to Do) with Time-Series Cross-Section Data," *American Political Science Review* 89 (September 1995). When we ran our regressions with the Beck-Katz procedure (estimating panel-corrected standard errors without instrumental variables), we obtained results that were essentially the same as those reported below (available upon request). None of our substantive findings are affected by the choice of one or the other of these setups.

[40] In terms of overall numbers, the interpolations make up for the loss of observations entailed by the use of a lagged dependent variable. More importantly, interpolation enables us to include Norway in our analysis.

V. Empirical Results

Table 4 reports the results of the linear regression model specified above and Table 5 reports the results of our interaction model. Table 6 pertains to the results of the interaction model, specifically, whether or not the observed differences between SMEs and LMEs are statistically significant.

Let us consider first the results of the linear model. While all four of the political-institutional variables in our analysis turn out to have a statistically significant association with levels of wage inequality, this is true for only one of the three variables intended to capture supply-and-demand conditions. The countervailing effects of unemployment, putting downward pressure on the wages of the unskilled workers but also removing a disproportionate number of unskilled workers from the employed labor force, apparently cancel each other out. As for LDC trade, the empirical results simply do not bear out our expectation that this variable would be positively associated with wage inequality. By contrast, female labor-force participation does have a sizable, positive coefficient, which readily clears conventional thresholds of statistical significance. This finding supports the hypothesis that increased female labor-force participation typically entails an increase in the relative supply of unskilled labor. As expected, union density, bargaining centralization, and government employment are associated with less wage inequality in the model, whereas rightist government is associated with more wage inequality. The size of the long-term coefficients for all these variables is considerable.

Setting aside the question of substantive significance for the time being, the results of our interaction model indicate that the determinants of wage inequality do indeed differ across clusters of advanced capitalist economies. Some of the generalizations supported by Table 4 appear quite misleading in view of the results reported in Table 5. To begin with, none of the variables in our analysis has statistically significant effects on the distribution of wages in the two countries we code as mixed cases (France and Italy). Thus we can concentrate on the theoretically interesting question of the differences between SMEs and LMEs.

Comparing SMEs with LMEs, we find that unemployment and LDC trade are equally irrelevant to wage-distributive outcomes while union density has a strong egalitarian impact in both sets of countries. More clearly than in the linear model, union density here emerges as the single most important variable explaining the observed variance of wage inequality in our data set. For the remaining four variables, we observe statistically significant differences between SMEs and LMEs (compare Table 6).

TABLE 4
THE DETERMINANTS OF WAGE INEQUALITY IN 16 COUNTRIES
(1973–95)

Variable	*Coefficients and Standard Errors*	*P-values*	*Long-Run Effects*
Lagged dependent variable	.499 (.153)	<.001	—
Unemployment	-.002 (.005)	.386	-.004
LDC trade	.0004 (.009)	.484	.0008
Female labor-force participation	.094 (.041)	.012	.188
Union density	-.037 (.018)	.020	-.074
Bargaining centralization	-.048 (.019)	.006	-.096
Government employment	-.099 (.042)	.009	-.198
Government partisanship	.022 (.007)	.001	.044

All entries are two-stage least squares estimates. P-values are one-sided. N = 217.

The coefficient for bargaining centralization is negative in both LMEs and SMEs, but the egalitarian impact of centralization is more than three times as great in SMEs according to these results. Indeed, the centralization coefficient for LMEs fails to clear the 90 percent level of statistical significance. On this count, our results diverge sharply from Wallerstein's. Whereas he writes that "it is difficult to find other variables that matter once the institutional variation in wage-setting is controlled for,"[41] we find not only that other political-institutional variables matter but also that the egalitarian effects of centralization are largely, if not entirely, contingent on SME conditions.[42]

The differences between SMEs and LMEs are more striking still with respect to the wage-distributive effects of female labor-force participation, government employment, and government partisanship, for the signs of the coefficients of these variables actually differ in the two sets of countries. According to our interaction model, the inegalitarian effects of female labor-force participation occur only under LME condi-

[41] Wallerstein (fn. 1), 650.

[42] Our analysis is based on a much larger number of observations and covers a longer time period than Wallerstein's. Different measures of bargaining centralization constitute another potential source of divergent results.

TABLE 5
THE DETERMINANTS OF WAGE INEQUALITY IN SOCIAL MARKET, LIBERAL MARKET, AND MIXED ECONOMIES (1973–95)

Variables	*Social Market Economies*			*Liberal Market Economies*			*Mixed Economies*		
	Coeffic. and Standard Errors	*P-values*	*Long-Run Effects*	*Coeffic. and Standard Errors*	*P-values*	*Long-Run Effects*	*Coeffic. and Standard Errors*	*P-values*	*Long-Run Effects*
Lagged dependent variable	.495 (.144)	*<.001*							
Unemployment	–.0004 (.006)	*.475*	–.0008	–.002 (.014)	*.433*	–.004	–.006 (.084)	*.472*	–.012
LDC trade	–.003 (.010)	*.368*	–.006	.013 (.015)	*.192*	.026	–.005 (.033)	*.442*	–.010
Female labor-force participation	–.059 (.071)	*.202*	–.117	.229 (.106)	*.016*	.453	–.374 (.641)	*.280*	–.741
Union density	–.114 (.048)	*.009*	–.226	–.119 (.046)	*.006*	–.236	–.060 (.105)	*.284*	–.119
Bargaining centralization	–.085 (.023)	*<.001*	–.168	–.024 (.021)	*.126*	–.048	–.024 (.051)	*.318*	–.048
Government employment	–.082 (.046)	*.040*	–.162	.104 (.052)	*.024*	.206	.006 (.508)	*.495*	.012
Government partisanship	–.005 (.011)	*.321*	–.010	.033 (.011)	*.002*	.065	.012 (.023)	*.305*	.024

All entries are two-stage least squares estimates. P-values are one-sided. N = 217.

TABLE 6
TESTS FOR EQUALITY OF COEFFICIENTS IN SMEs AND LMEs

Variable	*P-Value*
Unemployment	.898
LDC trade	.360
Female labor-force participation	.042
Union density	.934
Bargaining centralization	.005
Government employment	.018
Government partisanship	.015

P-values are for F-tests for the equality of the coefficients in SMEs and LMEs.

tions. The argument that social market conditions cushion the inegalitarian impact of the entry of new categories of unskilled labor into the workforce by extending union-negotiated wage contracts to nonunion workers provides a reasonable explanation of the absence of inegalitarian effects in the SMEs. However, this argument does not explain why female labor-force participation appears to be associated with less wage inequality in the social market economies of Western Europe, though the association is not quite significant by conventional criteria.

At least in part, the negative coefficient for female labor-force participation under social market conditions reflects the fact that the Scandinavian SMEs are characterized by higher rates of female participation as well as more compressed wages than are their continental counterparts. Arguably, the negative association between these variables is spurious, with female labor-force participation serving as a proxy for a complex of government policies that strengthens the position of women in the labor market and thereby reduces wage differentials between men and women. In other words, we might be picking up the effects of Scandinavian welfare states being more women friendly than the welfare states of continental Europe. But the negative association between female labor-force participation and wage inequality in SMEs might also be related to differences in the timing of increases in female labor-force participation. As noted earlier, the effects of female labor-force participation are likely to become egalitarian over time, as women acquire experience and skills at their jobs and as their position in the labor market becomes less precarious. As a group, the SMEs are more bifurcated on this score than the LMEs. Thus, the Scandinavian countries experienced a sharp increase of female labor-free participation as early as the 1960s, while female participation in the continental SMEs remained below the OECD average in the 1990s.

The results for government employment are similar to the results for female labor-force participation. In this case, the egalitarian effects observed in the linear regression model occur only among the SMEs. For the LMEs, the coefficient for government employment is positive—and statistically significant. The distinctively egalitarian dynamics of public sector wage bargaining appear to be contingent on SME conditions. Along the lines suggested above, our results are also consistent with the propositions that public sector wage premiums typically have inegalitarian effects and that the coordinated wage bargaining characteristic of social market economies curtail the scope of public sector wage premiums.

Finally, the results of our interaction model strongly support the hypothesis that SME conditions constrain the effects of government partisanship. In liberal market economies rightist government is associated with more wage inequality at a high level of statistical significance. In social market economies, by contrast, the association between these variables is far from significant.

The substantive significance of these findings is perhaps best illustrated by simulating the effects of a change in one or several of the statistically significant variables for a specific country. If we change the value of a particular variable to that of a country with more wage inequality, the simulated effect of this change can be compared with the actual difference in wage inequality between the two countries. Table 7 reports the results of simulations that assign German values to Sweden, using data for 1990 and the regression coefficients yielded by our interaction model. These two countries were chosen because Sweden, with a 90–10 ratio of 2.01 in 1990, stands out as the social market economy with the most compressed wage distribution, while Germany, with a 90–10 ratio of 2.72 in 1990, falls at the other end of the SME spectrum on our dependent variable.[41]

The first row of Table 7 contains our estimate of how the Swedish 90–10 ratio would have changed had Swedish union density suddenly dropped to the German level in 1990, with the values of all other variables remaining constant. Our regression results imply that such a change would have translated into a 24 percent increase of Swedish wage inequality over a ten-year period. In the second and third rows we report the results of repeating this exercise for the other two variables that proved to have a statistically significant association with wage inequality in SMEs, bargaining centralization, and government employ-

[41] In Figures 1 and 2 Austria has an even more dispersed wage distribution than Germany, but the fact that the Austrian wage data include part-time employees makes the Sweden-Germany comparison more appropriate.

TABLE 7
SIMULATED EFFECTS OF ASSIGNING GERMAN VALUES TO SWEDEN
(ACTUAL 90-10 RATIO IN 1990: 2.01)

			Simulation	
Variable[a]	*Swedish Level, 1990*	*German Level, 1990*	*90-10 Ratio after 10 Years*	*% Change in Wage Inequality*
Union density	84.0	32.9	2.48	+24
Bargaining centralization	.456	.318	2.13	+6
Government employment	31.6	15.1	2.26	+13
All 3 variables			2.97	+48

[a] When all three variables are changed at once, the total effect is realized in nine years, with the bulk of the effect realized in five years (90–10 ratio of 2.94 after five years).

ment. The fourth row, finally, contains our estimate of how Swedish wage inequality would have changed had all three variables taken on German values in 1990 (still holding all other variables in our analysis constant). When we change all three variables at once, Sweden's 90–10 ratio increases by 48 percent.

Similarly, Table 8 reports the results of simulations based on assigning U.S. values to Australia. As with Sweden and Germany, these countries were chosen because they represent opposite ends on the LME spectrum on our dependent variable. In 1990 the 90–10 ratio for Australia was 2.84, as compared with 4.33 for the U.S. Among LMEs there are four variables that have a statistically significant association with wage inequality: female labor-force participation, union density, government employment, and partisanship. Interpreting the results in Table 8 is complicated by the fact that as a pair Australia and the U.S. do not conform to our finding that there exists a positive relationship between the size of the public sector and wage inequality in LMEs. Assigning U.S. values to Australia on this variable actually reduces wage inequality in Australia (and increases the wage inequality gap between Australia and the U.S.). Nonetheless, changing the values of all four variables at once yields a 28 percent increase in wage inequality in Australia.

In a slightly different vein the British experience of the 1980s provides a convenient real-world illustration of the substantive significance of our estimates of the wage-distributive effects of government partisanship in liberal market economies. When the Conservatives took over the reins of government from the Labor Party in 1979, govern-

TABLE 8
SIMULATED EFFECTS OF ASSIGNING U.S. VALUES TO AUSTRALI
(ACTUAL 90–10 RATIO IN 1990: 2.84)

Variable[a]	Australian Level, 1990	U.S. Level, 1990	Simulation: 90–10 Ratio after 10 Years	Simulation: % Ch in W Inequ
Female labor-force participation	41.3	44.9	2.95	+4
Union density	46.6	15.6	3.68	+29
Government employment	23.0	14.6	2.59	–9
Government partisanship	2	4	2.97	+5
All 4 variables			3.64	+28

[a] When all four variables are changed at once, the total effect is realized in ten years, with the bul of the effect realized in five years (90–10 ratio of 3.61 after five years).

ment partisanship, as measured by Cusack's index, increased by 10C percent (from 2 to 4). The results of our interaction model imply that the results of this enduring change raised the British 90–10 ratio by about 5 percent (from 2.95 to 3.09) over the subsequent decade, but in fact, the British 90–10 ratio increased by 15.6 percent from 1979 to 1990. In a loose sense, our regression results thus suggest that almost one-third of the increase of British wage inequality might be attributed to Conservative government policies.

VI. CONCLUSION

While our analysis provides no support whatsoever for the contention that unemployment is a source of wage inequality or for the contention that trade with low-wage countries is a source of wage inequality, it lends some support to the hypothesis that female labor-force participation promotes wage inequality by increasing the relative supply of unskilled labor. However, this hypothesis holds only for countries characterized as liberal market economies. Should we conclude, then, that supply-and-demand conditions are largely irrelevant to the evolution of wage inequality in advanced capitalist economies? Our results certainly do not warrant such a far-reaching conclusion, which flies in the face of a great deal of empirical work (as well as theorizing) by labor economists. Variables such as the aggregate unemployment rate and

LDC trade as a percentage of GDP are probably too crude to capture the impact of market forces. At most, our analysis suggests that the wage-distributive effects of such variables are less straightforward than commonly supposed and that we should be wary of exaggerating their significance. In particular, the wage-distributive effects of trade with low-wage countries are likely smaller than suggested by some scholars (notably Wood) and by many media commentaries.[44]

Our findings concerning the role of political-institutional variables are less tentative and also more interesting. We have identified several discrete variables that fall under this general heading and we have shown that it is possible to obtain statistically significant results for all these variables in a pooled regression framework. We have shown furthermore that the distinction between social and liberal market economies has important implications for our understanding of the determinants of wage inequality. As with female-labor force participation, the wage-distributive effects of bargaining centralization, government employment, and partisanship differ across varieties of capitalism. Only one of the variables in our analysis, union density, has a significant association with wage inequality that is unaffected by the distinction between SMEs and LMEs.

The finding that the effects of partisanship are contingent on broad institutional constellations should be of particular interest to political scientists. Students of comparative politics have long argued over whether or not—or to what extent—the partisan composition of government matters to real economic and social outcomes, with the skeptics emphasizing that governments are constrained by some combination of political-institutional arrangements, structural economic conditions (specifically, the critical importance of private investment), and the preferences of the median voter. In recent literature on the evolution of government spending and social policy in advanced capitalist countries, this debate has been recast as a debate over whether or not partisanship still matters. While Pierson and Stephens, Huber, and Ray argue that partisan effects on spending have diminished since the 1970s, Garrett finds that these effects have actually become more pronounced in the era of globalization.[45]

[44] Wood (fn. 8). The more fine-grained analysis by Vincent Mahler, David Jesuit, and Douglas Roscoe also fails to establish any clear and consistent pattern of association between wage inequality and various dimensions of "globalization." See Mahler, Jesuit, and Roscoe, "Exploring the Impact of Trade and Investment on Income Inequality," *Comparative Political Studies* 32 (May 1999).

[45] Garrett (fn. 3); Paul Pierson, "The New Politics of the Welfare State," *World Politics* 48 (January 1996); John Stephens, Evelyne Huber, and Leonard Ray, "The Welfare State in Hard Times," in Kitschelt et al. (fn. 2); and Garrett (fn. 3).

Our analysis suggests that it may be useful to introduce the varieties-of-capitalism idea into the debate about partisanship: the extent to which partisanship matters and the extent to which partisan effects have changed over time may depend on how the political economy (as a whole) is organized. This said, we hasten to point out that the interaction effect we have established here pertains specifically to the determinants of wage inequality. The fact that partisanship matters to the distribution of wages in LMEs but not in SMEs reflects the broad scope of institutionalized wage bargaining and the concomitant absence (or insignificance) of minimum wage legislation and other forms of direct government regulation of wages in SMEs. It does not follow from this finding that the effects of partisanship in the realms of social policy or macroeconomic management are smaller in SMEs than in LMEs. Further research and analysis are clearly necessary to determine the broader implications of the varieties-of-capitalism approach for quantitative comparative political economy.

Needless to say, the results of our interaction model depend on our prior coding of countries as SMEs, LMEs, and MIXs. Most obviously, the decision to code Japan and Switzerland as liberal market economies might be questioned. By the criteria set out above, Japan and Switzerland certainly cannot be categorized as social market economies, but perhaps they should be categorized as mixed cases rather than liberal market economies. Focusing on public or collective regulation of labor markets, the picture presented in Table 2 entirely misses the informal practice of lifetime employment by large Japanese companies, for instance. Rerunning our interaction model with Japan and Switzerland coded as mixed cases (along with France and Italy), the regression results obtained for the mixed cases were quite different, but the results for the SMEs and the LMEs were essentially the same as those reported above. For the SMEs and the LMEs, no variable became either significant or insignificant at the 90 percent confidence level and differences in coefficient estimates between the SMEs and the LMEs that were initially significant at the 90 percent level remained so when we recoded Japan and Switzerland.[46] Practically speaking, then, our conclusions do not depend on whether Japan and Switzerland are considered liberal market or mixed economies.

[46] Among the statistically significant variables, the single largest coefficient change we obtained was for government employment in LMEs: this coefficient fell from .104 to .086 when we recoded Japan and Switzerland. For the mixed cases, the coding change produced statistically significant coefficients for wage bargaining centralization (negative, as in SMEs), government employment (negative, as in SMEs) and government partisanship (positive, as in LMEs). Results available from the authors upon request.

The time-invariant quality of the dummies we use to capture varieties of capitalism represents another potential pitfall. Perhaps recent changes in advanced capitalist political economies have rendered the SME-LME distinction less meaningful. Substantive knowledge of the country cases is the only real check on this problem. SMEs and LMEs alike underwent institutional changes during the time covered by our analysis (1973–95)—to some extent, captured by our union density and wage-bargaining centralization variables—but there is precious little evidence of generalized convergence between SMEs and LMEs and we simply cannot think of any country that can be said to have moved from the SME camp to the LME camp or vice versa.

More fundamentally, two alternative understandings of these clusters of advanced capitalist political economies should be noted here. Consider the matter in terms of our dummy variable for SMEs. On the one hand, this variable might be viewed as a proxy for a set of discrete variables—welfare-state decommodification, institutionalization of collective bargaining, and employment protection—for which we do not have year-to-year observations. Some of these variables may be difficult to measure, but in principle they are all measurable, and if we had observations for each of these variables, the SME dummy would be superfluous. On the other hand, we might think of the SME dummy as capturing the way that discrete variables are configured within a coherent whole, that is, as referring to an institutional context with emergent properties that cannot be reduced to discrete variables. Setting aside the specificities of the SME concept, the tension between these two perspectives animates current theoretical debates in comparative political economy. While we are inclined to take the idea of emergent properties seriously, we want to emphasize the empirical nature of the questions at stake in this debate. The most obvious way to determine whether varieties of capitalism are reducible to discrete variables is to try to achieve such a reduction.

Finally, the importance of union density as a determinant of wage inequality deserves to be underscored one more time. Across the SME-LME divide, the effects of union density are consistently egalitarian, and for each cluster of countries they are greater than those of any other independent variable in our analysis. This finding suggests that conflicts of interest between unions and employers constitute an important dimension of the politics of wage distribution. If wage-distributive outcomes were primarily an expression of employer preferences, as Swenson's revisionist account of the Swedish story of wage solidarity would have it, it would be difficult to make sense of the association between

union density and wage compression.[47] On this score, the general impli cation of our analysis is that we should be wary of embracing a wholl employer-centered approach to comparative political economy.

In no way does our analysis deny that distributive conflict amon wage earners also constitutes an important dimension of the politics o wage distribution. The empirical data at our disposal simply do no allow us to capture this dimension. To do so, we would need to be abl to measure gender differences in union density or the distribution c union membership across the hierarchy of wages and skills. (By shiftin the focus to intersectoral wage differentials, the sectoral distribution o union membership would also be of interest.) The labor force survey from which such measures might be derived are becoming availabl but only for recent years. They thus do not allow very sophisticate analyses using countries as the unit of observation. Tackling thi methodological challenge is an important part of the research agend that flows from the preceding analysis. In future research we also wan to explore the interface between distributive labor-market outcome and the redistributive effects of the welfare state.

APPENDIX: DATA SOURCES

90–10 ratio. data provided by the OECD, Directorate for Education Employment, Labour and Social Affairs.

Unemployment. OECD, *Historical Statistics* (electronic database).

LDC trade. For all countries but Belgium, data on LDC trade were pro vided by Geoffrey Garrett (Yale University); Belgian and post-199 figures were calculated on the basis of OECD, *Monthly Statistics of For eign Trade.*

Female labor-force participation. OECD, *Historical Statistics* (electroni database).

Government employment. OECD, *Historical Statistics* (electronic data base).

Union density. The pre-1990 figures were taken from Jelle Visse "Unionization Trends Revisited," Centre for Research of European So cieties and Industrial Relations (Amsterdam, 1996); post-1990 figure were provided by Bernhard Ebbinghaus (Max-Planck Institute).

[47] Peter Swenson, *Labor Markets and Welfare States* (forthcoming).

Centralization. The wage-bargaining centralization index was provided by Torben Iversen (Harvard University). For a complete specification, see Iversen (fn. 3).

Government partisanship. Figures were provided by Tom Cusack (Wissenschaftszentrum, Berlin). For details, see Cusack (fn. 24).

Research Note

REASSESSING THE THREE WAVES OF DEMOCRATIZATION

By RENSKE DOORENSPLEET*

INTRODUCTION

SINCE the publication of Samuel Huntington's influential 1991 study of democratization,[1] scholars have come to take for granted the notion that the spread of democracy has come in waves, with bursts of progress being succeeded by quite substantial reversals—the result being a flow and ebb marking a less than optimistic two-step-forward, one-step-backward pattern. According to Huntington, there have in fact been three distinct waves of democratization, with a wave being defined as a group of transitions from nondemocratic to democratic regimes that occurs within a specified period of time and in which those transitions significantly outnumber transitions in the opposite direction.[2] This is far from a linear process, however, as waves of democratization have been followed by reverse waves in which some of the democratic countries reverted to nondemocratic rule, leaving fewer cases of consolidated democracies behind. According to Huntington, the first "long" wave flowed uninterruptedly from 1826 to 1926, marking the emergence of democratic regimes as a nineteenth-century phenomenon. Following a reverse wave, the end of World War II provided the impetus for the second, short wave of democratization. Thereafter came an enormous global swing away from democracy in the 1960s and early 1970s, which, in turn, was succeeded by a third wave of democratization, which took off in the years following the end of the Por-

* Special thanks to Peter Mair, whose critical comments and editorial assistance were indispensable to this article. I am also indebted to Ed Greenberg, Wil Hout, John O'Loughlin, Hans Oversloot, Huib Pellikaan, and Mike Ward for their careful readings and encouragement. This work was supported by the Foundation for Law and Public Administration (Reob), which is part of the Netherlands Organization for Scientific Research (NWO).

[1] Samuel P. Huntington, *The Third Wave: Democratization in the Late Twentieth Century* (Norman: University of Oklahoma Press, 1991).

[2] Ibid., 15.

tuguese dictatorship in 1974.[3] Arguing from the vantage of the early 1990s, Huntington further contended that these waves have entailed a far from optimistic pattern of progress:

> The proportions of democratic states in the world show a considerable regularity. At the troughs of the two reverse waves 19.7 percent and 24.6 percent of the countries were democratic. At the peaks of the two democratization waves, 45.3 percent and 32.4 percent of the countries in the world were democratic. In 1990 roughly 45.4 percent of the independent countries of the world had democratic systems, the same percentage as in 1922. . . . [I]n 1990 the third wave of democratization still had not increased the proportion of democratic states in the world above its previous peak sixty-eight years earlier.[4]

Moreover, he also suggested that, at the time he was writing, there were possible signs of the beginnings of a third reverse wave, in that three third-wave democracies—that is, Haiti, Sudan, and Surinam—had quickly reverted to authoritarianism.[5]

Although Huntington's study has been very influential, I wish to suggest that his analysis is nevertheless far from compelling. There are two problems. The first is largely conceptual: the analysis fails to provide a clear and meaningful distinction between democratic and authoritarian regimes. To be sure, Huntington's definition of democracy adopts Dahl in specifying the three requirements of competition, inclusiveness, and civil liberties.[6] But in practice he focuses primarily on Dahl's dimension of competition and sometimes simply ignores the requirement of universal suffrage—that is, the equally important dimension of inclusion.[7] Thus, Huntington's eventual classifications are more ambiguous and inconsistent, as for instance when he classifies the

[3] By Schmitter's reckoning, there have been four, more compact waves. In addition to Huntington's second and third waves, he thinks there had occurred two other earlier waves; one spectacular but ephemeral wave began in 1848 and reverted in 1852, and the other major outbreak of democracy corresponded to World War I and its aftermath. See Philippe C. Schmitter, "Waves of Democratization," in Seymour Martin Lipset et al., eds., *Encyclopedia of Democracy*, vol. 2 (London: Routledge, 1995), 346–50.

[4] Huntington (fn. 1), 25–26.

[5] Ibid., 14–15.

[6] Robert A. Dahl, *Polyarchy, Participation and Opposition* (New Haven: Yale University Press, 1971); Huntington (fn. 1), 7.

[7] Huntington is not alone in this. Many other quantitative researchers focus almost exclusively on the degree of competition and make little or no reference to the extent to which the regimes in question are also inclusive. See, e.g., Kenneth A. Bollen, "Issues in the Comparative Measurement of Political Democracy," *American Sociological Review* 45 (June 1980); idem, "Liberal Democracy: Validity and Method Factors in Cross-National Measures," *American Journal of Political Science* 37 (October 1993); Raymond D. Gastil, "The Comparative Survey of Freedom: Experiences and Suggestions," in Alex Inkeles, ed., *On Measuring Democracy: Its Consequences and Concomitants* (New Brunswick, N.J.: Transaction Publishers, 1991); Keith Jaggers and Ted Robert Gurr, "Tracking Democracy's Third Wave with the Polity III Data," *Journal of Peace Research* 32 (1995); M. Alvarez et al., "Classifying Political Regimes," *Studies in International Comparative Development* 31 (Summer 1996).

United States until 1965 and Switzerland until 1971 as undemocratic, while some pages later indicating that both were democratic a century earlier.[8] Moreover, he also appears to adopt other criteria for nineteenth-century systems, which are classified as having already become democratic when 50 percent of adult males become eligible to vote.[9] Finally, early-twentieth-century Portugal is also considered democratic, even though only male citizens had then won the right to participate.

The second problem, equally acute, is empirical. In brief, Huntington has estimated the incidence of transitions to democracy in terms of the *percentages* of world states involved. Since the denominator in this equation, that is, the number of states in the world, is far from constant, this measure can be misleading. As we shall see, for example, the number of (minimal) democracies in the world grew from thirty in 1957 to thirty-seven in 1972, thus appearing to reflect a small but noticeable "wave" of democratization. Considered as a percentage, by contrast, this same period seems to have been characterized by a small *reverse* wave, in that the proportion of states that were democratic fell from 32 percent to 27 percent. The explanation for this apparent paradox is simple: largely as a result of decolonization in Africa, the number of independent states in the world—the denominator—grew from 93 to 137; hence, although there was an absolute increase in the number of democratic regimes, their proportion of world states actually fell.

In the subsequent sections of this research note I offer solutions to both of these problems, first, by specifying a tighter and more consistent distinction between democratic and nondemocratic regimes, and second, by taking into account the actual numbers of states that made a transition from nondemocracy to democracy, or vice versa. I conclude by proposing a reassessment of the three waves of democracy that suggests that there has been a first and major wave of democratization lasting from the early 1890s to the mid-1920s. This is then followed by a minor reverse wave that continues through the early 1940s, after which the trend turns upward again, although not in so pronounced a fashion as that suggested by Huntington's data. Strikingly, no second reverse wave is really apparent. Rather, the period from the late 1950s to the mid-1970s may be characterized as one of relatively trendless fluctuation, leading to the breaking of a new wave of democratization in 1976 and the subsequent explosion of democratization since 1990.

[8] Huntington (fn. 1) 7, 14–17.
[9] Ibid., 16.

These findings are important not only because they suggest a somewhat altered sequencing and a more accurate count of democratization waves but also because they cast real doubt on the appropriateness of the wave metaphor. With regard to the spread of democracy we certainly see flows; the ebbs, however, are less evident, indeed, are not really apparent at all from these data. Future studies should therefore also be cautious in comparing, explaining, and forecasting different waves of democratization.

THE PROBLEM OF CLASSIFICATION AND CONCEPTUALIZATION

Following Dahl, [10] Huntington defines democracy as a political system that meets the three requirements of competition, inclusiveness, and civil liberties.[11] This definition, which basically corresponds to more recent attempts to define liberal democracy, has become increasingly important as scholars have drawn attention to the increasing number of countries that are characterized by inclusive suffrage and open competition but that at the same time lack a fully developed system of civil liberties.[12] These countries underline the fact that the dimension of civil liberties can be quite independent of the dimensions of competition and inclusion. In keeping with the more recent literature, these are defined here as "minimal democracies."[13]

Although Huntington's definition is clear in seeming to focus on liberal democracies, his actual classification of political regimes is not transparent.[14] It is not at all clear, for example, how the dimension of

[10] Dahl (fn. 6) emphasizes that there is no country in which these conditions are perfectly met; therefore he prefers the term "polyarchies" for political systems in which the conditions are sufficiently met and uses the term "democracy" for the ideal type. In this research the term "liberal democracy" will be used for Dahl's polyarchies, because the term "democracy" is more common in daily language.

[11] Huntington (fn. 1), 7.

[12] Cf. Larry Diamond, Juan Linz, and Seymour Martin Lipset, *Politics in Developing Countries: Comparing Experiences with Democracy* (Boulder, Colo.: Lynne Rienner, 1995); Larry Diamond, "Is the Third Wave Over?" *Journal of Democracy* 7 (July 1996); Mark J. Gasiorowski, "An Overview of the Political Regime Change Dataset," *Comparative Political Studies* 29 (August 1996); Andreas Schedler, "What Is Democratic Consolidation?" *Journal of Democracy* 9 (April 1998); Schmitter (fn. 3); George Sorensen, *Democracy and Democratization* (Boulder, Colo.: Westview Press, 1998); Fareed Zakaria, "The Rise of Illiberal Democracy," *Foreign Affairs* 76 (November–December 1997).

[13] Multiple labels are applied to indicate the same concept; this type of regime has been variously described. It is called "formal" or "electoral democracy" in Axel Hadenius, "The Duration of Democracy: Institutional versus Socio-Economic Factors," in David Beetham, ed., *Defining and Measuring Democracy* (London: Sage Publications, 1994), 69; and in Diamond (fn. 12). It is called "democradura" in O'Donnell and Schmitter, "Tentative Conclusions about Uncertain Democracies," in O'Donnell and Schmitter, eds., *Transitions from Authoritarian Rule: Prospects for Democracy* (Baltimore: Johns Hopkins University Press, 1986), 9. And it is called "illiberal democracy" in Zakaria (fn. 12).

[14] Dankwart A. Rustow, "The Surging Tide of Democracy," *Journal of Democracy* 3 (January 1992), 121.

civil liberties is measured. In addition, what is sometimes neglected in Huntington's treatment is that account be taken of *both* of Dahl's other two dimensions, that is, universal suffrage—or inclusiveness—in addition to competition. Although it is a commonplace to incorporate the inclusiveness dimension into the *concept* of democracy, this dimension is often ignored in the actual *measurement* of democracy. It is therefore still necessary to underline the importance of this dimension. Indeed, it is surprising how many quantitative studies, like Huntington's, ignore the suffrage dimension and hence consider more exclusive regimes as also being "democratic."[15] Bollen, for example, neglects this dimension almost entirely, arguing that (voter) participation is only marginally related to democracy, particularly because voter participation also has a symbolic value that is, or was, often employed in nondemocratic countries. Thus he cites data from the early 1970s to show that Albania, North Korea, the Soviet Union, Romania, and Bulgaria were among those countries recording the highest levels of electoral participation while noting that "very few researchers would consider any of these countries as highly democratic."[16] Further evidence that Bollen cites against using electoral participation as a valid measure of political democracy is its low level of correlation with other attested indicators of democracy, such as the fairness of elections.

There is a danger here, however. By neglecting the inclusiveness dimension, one risks adopting a measurement of democracy that is biased or even racist or sexist. Bollen, to illustrate, will consider a regime with extensive opportunities for liberalization as democratic, even though only a very small proportion of the society (for example, only white men) may be allowed to participate.[17] Further, although it is obviously true that electoral participation may be of symbolic value to authoritarian regimes, which cannot be considered democratic, this is hardly a compelling argument for excluding the dimension of inclusiveness altogether. In Dahl's initial formulation, such regimes were defined as inclusive hegemonies, that is, as inclusive regimes without competition. Finally, the fact that these two dimensions scarcely correlate does not constitute an argument against incorporating inclusiveness as a property of democracy; rather, this simply indicates that these are in fact

[15] E.g., Bollen (fn. 7, 1980 and 1993); Gastil (fn. 7); Jaggers and Gurr (fn. 7); Alvarez et al. (fn. 7).
[16] Bollen (fn. 7, 1980), 373.
[17] Likewise, in Gurr's Polity III data set the South African Apartheid regime gets a score of 5. This score is quite high on the scale from −10 to +10, especially if one bears in mind that the Fifth Republic of France gets a score of 6. The inclusiveness dimension is clearly ignored in Gurr's measurement.

two quite independent dimensions. For the purposes of this research note, therefore, regimes in which competition coexists with noninclusive suffrage—Dahl's competitive oligarchies—will be regarded as nondemocratic regimes.[18]

What is required, therefore, is a classification that not only is transparent and consistent but that also incorporates inclusiveness. In fact, four main types of regime may be distinguished here: liberal democracies, minimal democracies, authoritarian systems, and interrupted regimes. A liberal democracy is a regime in which there is meaningful and extensive competition, sufficiently inclusive suffrage in national elections, and a high level of civil and political liberties. Minimal democracies are those political regimes with competition and inclusive suffrage but without a high level of civil liberties. Authoritarian regimes are those political regimes that fail to meet the first requirement of competition and/or the second requirement of inclusiveness. Finally, a country will be classified here as an interrupted regime if it is occupied by foreign powers during wartime, or if there is a complete collapse of central authority, or if it undergoes a period of transition during which new polities and institutions are planned.[19]

Not all of these criteria are equally susceptible to long-term historical analysis, of course, and a lack of comparable evidence over time makes it impossible to measure Dahl's third requirement of civil liberties.[20] Accordingly, this analysis does not include the waves of liberal democracies and is limited to the establishment and withdrawal of the minimal democratic systems.

[18] It should also be pointed out that voter turnout is not an adequate indicator of inclusiveness, reflecting instead factors that have little to do with measuring the right to participate in national elections (inclusiveness). Rather, structural indicators such as the institutional guarantees to participate in elections are required to calculate this second requirement of democracy. This also means that the many rich data sets on voter turnout cannot be used in this research. These include, for example, Arthur S. Banks, *Cross-Polity Time-Series Data*, assembled by Arthur S. Banks and the staff of the Center for Comparative Political Research, State University of New York at Binghamton (Cambridge: MIT Press, 1971); Tatu Vanhanen, *The Emergence of Democracy: A Comparative Study of 119 States, 1850–1979* (Helsinki: Finnish Society of Sciences and Letters, 1984); idem, *Prospects of Democracy: A Study of 172 Countries* (London: Routledge, 1997); IDEA, *Voter Turnout from 1945 to 1997: A Global Report* (Stockholm: IDEA, 1997).

[19] There are other quite equivalent and interesting classifications of democracy, but these are not useful for the present research. See Mitchell Coppedge and Wolfgang H. Reinecke, "Measuring Polyarchy," in Inkeles (fn. 7); Alvarez et al. (fn. 7); Gasiorowski (fn. 12) These classifications are available only for one specific year (e.g., Coppedge and Reinecke), or for a limited period since 1950 (Alvarez et al. [fn. 7]), or they are limited to developing countries (Gasiorowski [fn. 12]). What are needed in the context of this study are measures of democracy that cover both a long period of time and all independent countries.

[20] Gastil's data on civil liberties are available only for the period since 1973, and hence transitions toward liberal democracies can be determined and investigated only since then.

MEASURING MINIMAL DEMOCRACY

The first requirement of minimal democracies, the presence of competition, can be seen to be met if there exist institutions and procedures through which citizens can express effective preferences about alternative policies at the national level and if there are institutionalized constraints on the exercise of power by the executive. Indicators of these phenomena have already been collated in Gurr's well-known Polity III data set, which covers most independent countries on an annual basis from 1800 to 1994; it is therefore an ideal source to measure the presence of competition.[21] Moreover, these data are also easily adapted to the definition of competition employed in this analysis.[22] In operational terms, I will consider a national political system to be competitive if there is at least one executive chosen by competitive popular elections (if Gurr's indicator "competitiveness of executive recruitment" is coded 3 or 4); if all the politically active population has an opportunity, in principle, to attain an executive position through a regularized process (if Gurr's indicator "openness of executive recruitment" is coded 3 or 4); if alternative preferences for policy and leadership can be pursued in the political arena, such that oppositional activity is not restricted or suppressed (if Gurr's indicator "competitiveness of participation" is coded 0, 3, 4, 5); and if there are at least substantial limitations on the exercise of executive power (if Gurr's indicator "constraints on the power of the chief executive" is coded 4, 5, 6, or 7).[23]

The second requirement of minimal democracies is that there be inclusive, universal suffrage at the national level. The norm of universality requires that all citizens of the state—without regard to sex, race, language, descent, income, land holdings, education, or religious beliefs—formally enjoy the right to vote and to be elected to public office. The fact that certain prerequisites are demanded, such as a minimum age, a sound mind, or the absence of criminal record, does not negate this principle. Only countries that at some stage meet the first requirement of competition from 1800 to 1994 are considered when measur-

[21] Ted Robert Gurr, "Persistence and Change in Political Systems," *American Political Science Review* 74 (December 1974); Ted Robert Gurr, Keith Jaggers, and Will H. Moore, "Polity II Codebook" (Manuscript, University of Colorado, Boulder, 1989); Gurr, Jaggers, and Moore, "The Transformation of the Western State: The Growth of Democracy, Autocracy, and State Power since 1800," in Inkeles (fn. 7); Jaggers and Gurr (fn. 7).

[22] Gurr himself talks about scores of autocracy and democracy, but it is nevertheless clear that what he is measuring is competition. That is why I label this scale a "competition-scale."

[23] These choices are more fully explained in Renske Doorenspleet, "Democracy, Transitions, and Waves" (Manuscript, Department of Political Science, University of Leiden, the Netherlands, July 1998).

ing the inclusiveness of the system.[24] Levels of inclusiveness of the political system may be broken down into one of the following four categories: (1) no popular suffrage; (2) suffrage denied to large segments of the population (more than 20 percent is excluded); (3) suffrage with partial restrictions (less than 20 percent of the population is excluded); (4) universal suffrage or minor restrictions.[25] For the purposes of this analysis, countries are considered "sufficiently" inclusive to meet the criterion of a minimal democracy if they fall into the third or fourth category.[26] Should they fall within either of the first two categories or should they not meet the competition criterion, they are classified as authoritarian regimes. Reliable data on inclusiveness are of course difficult to obtain and to standardize, and I have had to rely on historical sources and various monographs for each country, as well as on *Keesing's Record of World Events* and many of the standard handbooks and almanacs.[27] The appendix gives an overview of the years in which political systems can be considered as both competitive and inclusive and hence are classified in this study as "minimal democracies."

Interrupted regimes are defined as those countries that are occupied by foreign powers during wartime (for example, the Netherlands during

[24] Countries that did not pass the requirement of competition are authoritarian; therefore, in the context of this research there is no need to investigate the inclusiveness of these noncompetitive systems.

[25] Coppedge and Reinecke (fn. 19) used the same four categories in measuring inclusive suffrage for all political regimes in 1985. The authors dropped this variable (and dimension) of inclusiveness from the final scale because they came to the conclusion that it contributed very little empirically to the measurement of polyarchy. Consequently, their final scale is a unidimensional scale of contestation. This result is hardly surprising, in that by 1985 the dimension of "inclusiveness" had largely ceased to play an independent effect. In earlier stages of democratization, on the other hand, it is of crucial importance.

[26] This is in fact a critical choice: when do countries fulfill the requirement of inclusiveness sufficiently? Of course, if a country falls in the first category. But what if there are partial restrictions? In the ent, however, only two countries appear to fall in the second category "suffrage with partial restrictions" (less than 20 percent of the population is excluded): Chile excluded 10 percent of the population literacy requirements until 1970, and in the United States, a variety of devices, such as poll taxes, literacy tests, and property qualifications, prevented virtually all blacks (10 percent of the population) om voting. I decided to treat these two cases as exceptional ones and I will consider Chile in 1949 d the United States in 1920 as "inclusive" enough when they granted women the right to vote. It is to be pointed out, in addition, that a country is considered to be inclusive *since* the year in which e formal right to participate can be carried out, that is, since the year in which the first inclusive elections are actually held. Hence, although Dutch women had the formal right to vote already in 1919, ey could not exercise this right until the 1922 elections; consequently, the Netherlands is classified as "minimal democracy" only since 1922.

[27] Examples of the handbooks and almanacs that were used to investigate the inclusiveness over time the political regimes in this study are Chris Cook and John Paxton, *European Political Facts, 18–1973* (London: MacMillan, 1975); George E. Delury et al., *World Encyclopedia of Political Systems*, 2 vols. (Essex: Longman, 1983); Ian Gorvin et al., *Elections since 1945: A World-wide Reference ompendium* (Harlow: Longman, 1989); Lipset et al., eds., *The Encyclopedia of Democracy*, vols. 1–4 ondon: Routledge, 1995); Thomas T. Mackie and Richard Rose, *The International Almanac of Electoral History* (London: MacMillan, 1991); Stein Rokkan and Jean Meyrat, *International Guide to Electoral Statistics* (The Hague and Paris: Mouton, 1969).

World War II), that experience a complete collapse of central authority, say, during civil war (for example, Lebanon, 1978–86), or that undergo a period of transition during which new polities and institutions are planned, legally constituted, and put into effect (for example, Argentina in 1982). Gurr's Polity III data set includes useful data recording these experiences, and in operational terms, I have chosen to classify countries as interrupted when they have Gurr codes of 66, 77, or 88.[28]

Waves of Democracy

Having established these criteria, I now consider how the waves of democracy developed over time. First, and following researchers who measure democracy by taking account only of Dahl's requirement of competition, the empirical trends in the growth of "democratic" systems from 1800 to 1994, measured as a percentage of all countries, are summarized in Figure 1. This appears more or less to replicate Huntington's findings, albeit here updated to 1994. From these figures, his three waves of democratization can in fact be readily distinguished. In addition, Figure 1 shows that the staging points of democratization in terms of waves and reverses generally closely resemble those identified by Huntington: there is a first, long wave of democratization (1810–1922), a first reverse wave (1923–40), a second wave of democratization (1944–57), a second reverse wave (1957–73), and a third wave of democratization (since 1973).

Although these updated figures fail to confirm Huntington's statement that there is a less than optimistic two-step-forward, one-step-backward pattern, they nevertheless do show that the proportion of democratic states during the third wave rises well above its previous peaks. Huntington's less optimistic conclusions are probably simply due to the fact that his analysis ended in 1990, whereas in the following few years the proportion of democratic states increased enormously, from 43 percent in 1990 to more than 57 percent in 1994. This percentage is much higher than the percentages of democracies in the first peak; that is, it is higher than the 38 percent in 1922, and it is also higher than the 35 percent in 1960. So although the trend toward democratization as defined in these limited terms may not be irreversible, the long-term trend in Figure 1 does not appear to reflect a genuine two-step-forward, one-step-backward pattern and reveals no signs as yet (through to 1994) to indicate a third reverse wave.

[28] Gurr, Jaggers, and Moore (fn. 21, 1989), 6–8.
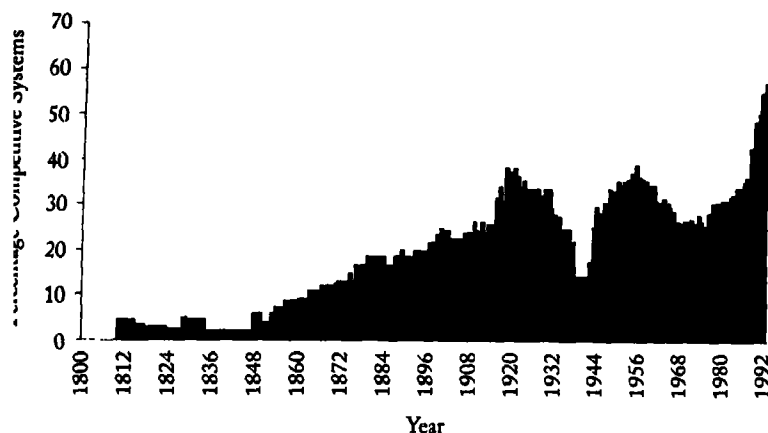
FIGURE 1
PERCENTAGE COMPETITIVE SYSTEMS
(BY YEAR)

WAVES OF MINIMAL DEMOCRACY

As is evident, however, these data fail to include the requirement of inclusive suffrage and therefore do not refer to the development of what we define as minimal democracies. Once this second measure is included for this more complete definition of minimal democracy, we may expect to see a very different pattern. As can be seen from Figure 2, which plots the percentage of minimal democracies from 1800 to 1994, this is in fact the case.

In this new application, the first wave can be seen to begin at a much later stage, confirming that democratization in this more complete sense (of taking among others women's citizenship seriously) is a twentieth-century phenomenon that develops also relatively abruptly. While there were as yet no minimal democracies as late as 1890, by 1923 almost a quarter of the independent political systems in the world had achieved that status. The first peak is lower in Figure 2 than in Figure 1, in that several competitive systems had not yet extended the suffrage to women (for example, Belgium, Costa Rica, France, Portugal, Spain, Switzerland) and hence, strictly speaking, cannot be considered as minimal democracies. The second peak, in 1955, is also slightly lower (by 4.5 percent) in Figure 2, in that Brazil, Peru, Sudan and Switzerland, while competitive, were not yet minimal democracies. The second reverse wave, which is now much smaller in comparison with Figure 1,

FIGURE 2
PERCENTAGE MINIMAL DEMOCRACIES
(BY YEAR)

results in a fall to 27 percent in 1973. However it should also be noted that from this point onward the percentages are the same in both figures, thus indicating that the requirement of inclusive suffrage no longer exerts any influence. Hence the third wave is also evident in Figure 2.

EMPIRICAL PROBLEMS

Until now, I have followed the researchers who studied the waves of transitions by focusing on percentages of democracies among world states.[29] But does such a focus really offer insight into the waves of transitions? Remember that a wave of democratic transitions is defined as a group of transitions from authoritarian regimes to (minimal) democracies that occurs within a specified period of time and in which the transitions significantly outnumber transitions in the opposite direction. In investigating such waves, the focus on percentages of democracies in the world can be very confusing, for two reasons in particular.

First, although there may be fewer minimal democracies during a reverse wave, this does not necessarily mean that those states have reverted to authoritarianism; it may mean, instead, that several states have experienced an interrupted period. Consider, for example, the first reverse

[29] Huntington (fn. 1); Jaggers and Gurr (fn. 7); Kristian S. Gleditsch and Michael D. Ward, "Double Take: A Reexamination of Democracy and Autocracy in Modern Polities," *Journal of Conflict Resolution* 41 (June 1997).

wave from 1923 to 1940, which seems to be evident in Figure 2. Indeed, there were certain democratic regimes, such as Poland, Germany, and Italy, which did retreat to authoritarianism during this period. There were, however, also other countries, such as Luxembourg, the Netherlands, Norway, and Denmark, that were minimal democracies in the 1920s and yet did not really step backward. What actually happened was that these countries experienced a difficult period of interruption following the German invasion and the subsequent foreign domination of their own national political regimes before being restored to democracy in 1945. In neither case, however, should they be considered to be part of a wave or reverse wave, in that they neither moved toward (in 1940) nor away from (in 1945) authoritarianism. While Huntington does not distinguish between such transitional regimes and genuinely authoritarian regimes (he simply considered transitional countries to be authoritarian), I believe this distinction to be particularly important. Thus, to gain a more accurate sense of (reverse) waves of democratization, it would seem more appropriate to exclude these interrupted cases and hence to plot only the genuine percentages of democracies and authoritarian regimes. The principal effect of doing this, of course, is to reduce the levels of the first reverse wave of transitions.

A second and more crucial empirical problem is that changes in the proportion of minimal democracies around the world can occur simply because the denominator (the number of states) itself changes. As can be seen from Figure 2, for example, the percentage of minimal democracies in the world falls from more than 32 percent in 1957 to 26 percent in 1972—an evident reversal, or so it would seem. Yet at the same time the sheer numbers of minimal democracies actually increased from thirty to thirty-seven, suggesting significant progress toward democracy. In fact, this apparent paradox is easily explained by reference to the widespread decolonization in Africa and the subsequent enormous growth of independent states from 93 in 1957 to 137 in 1972. Moreover, most of these new states were authoritarian—in 1972 only eight of the forty-four new states were democratic. Huntington considers this as part of a transition process, suggesting that these new states *became* authoritarian at independence.[30] In my view, however, this is certainly not evidence of a "global swing away from democracy"[31] but is rather a process by which new authoritarian states were simply created or installed. Hence the second reverse wave that appears in Figure

[30] Huntington (fn. 1), 20–21.
[31] Ibid., 21.

2 should not be interpreted as exclusively reflecting regime transitions away from democracy; it also reflects the *installation* of many newly independent authoritarian regimes.

From this it follows that we can gain a better and more accurate insight into the real *transitions* toward (and away from) minimal democracy and hence into the waves of democracy by excluding those states that have either (1) experienced an interruption of their own national regimes or (2) become newly established as independent regimes,[32] and by focusing on the real numbers rather than on the otherwise misleading percentages. In other words, by plotting the number of transitions from (independent) authoritarian to democratic systems less the number of transitions from minimal democracies to authoritarian systems over time, we gain a more accurate and meaningful insight into the question of whether the transitions to democracy outnumber the transitions to authoritarianism or vice versa. In 1920, for example, this difference is (+) 5, in that Austria, Czechoslovakia, Denmark, Estonia, and the United States each became a minimal democracy, while no single country followed the reverse path. In 1973, by contrast, the difference is (+) 1, in that two countries became democratic (Argentina and Turkey), while one other reverted to authoritarianism (Chile). The overall results of this revised analysis are summarized in Figure 3.

In order to effect a more appropriate comparison over time, however, it may still be necessary to build in some control for the growth in the number of states in the world, particularly since this may increase not only the probability that transitions will take place but also the difference between transitions to minimal democracy and transitions to authoritarianism. Consider the following hypothetical situations during two points in time: (1) there are 12 countries, of which 3 make a transition to authoritarianism and 4 to democracy; (2) there are 120 countries, of which 30 make a transition to authoritarianism and 40 to democracy. In the first situation, transitions to democracy outnumber transitions in the opposite direction by 1, while in the second situation the relevant figure is 10. In terms of relative proportions, however, the situations are identical, and hence each might be seen to reflect an equivalent "wave." Comparisons over time may therefore require the data to be standardized in the following manner:

[32] One could object that I neglect the new state's choice of its own regime type by excluding those states that have become newly established as independent regimes. However, Huntington himself defines a wave as a group of transitions, and I do not regard as "transitions" the institution of new regimes
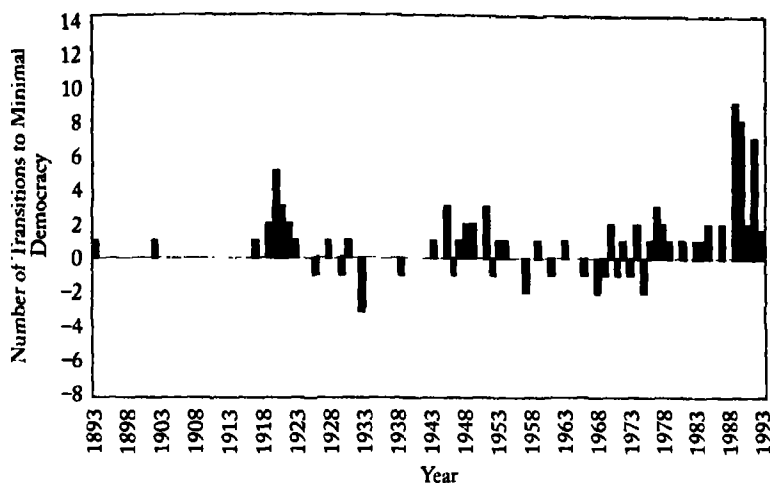
FIGURE 3
NUMBER OF TRANSITIONS TO MINIMAL DEMOCRACIES
(BY YEAR)

WT (*t*) = N (1994) * Trans (*t*) / N (*t*)
where
WT (*t*) = weighted number of transitions in year *t*
N (1994) = number of countries in 1994 (= 156)
Trans (*t*) = number of transitions in year *t*
N (*t*) = number of countries in year *t*

This formula takes the year 1994 as the standard against which the number of transitions will be weighted in this study. (See the appendix for the list of regime changes from 1800 through 1994.) The standardized data are summarized in Figure 4, with more detailed figures on the different waves indicated there being presented in Table 1.

Based on the standardized data and with a view to reassessing Huntington's three waves, the following periods can be distinguished: a first impressive wave of democratization in 1893–1924; a first reverse wave in 1924–44; a second wave of democratization in 1944–57; a period of relatively trendless fluctuation in 1957–76; and finally a third wave of democratization in 1976–89 that is then followed by the explosion of democratization since 1990. The first wave of transitions to democracy till 1924 is clearly very important and very striking. It is also clear that there has been a significant wave of democratization since 1976 and
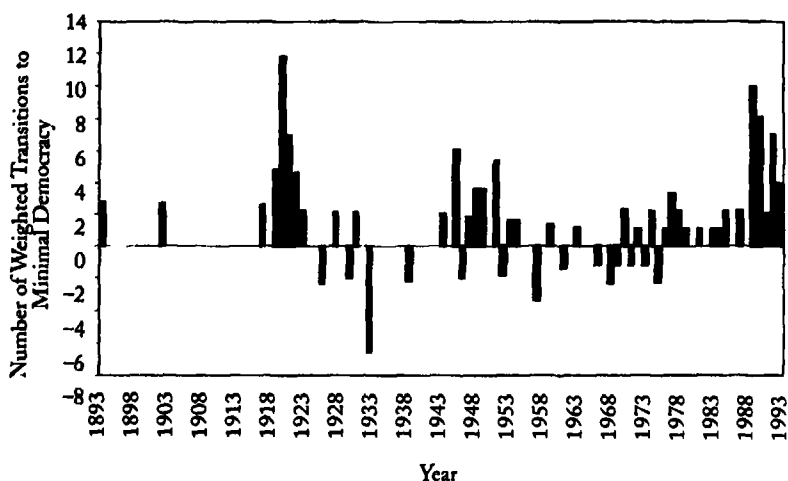
FIGURE 4
NUMBER OF WEIGHTED TRANSITIONS TO MINIMAL DEMOCRACIES
(BY YEAR)

that there has been a real upsurge of transitions to democracy since 1990. What is perhaps more striking, however, is that Huntington's other waves no longer emerge with such significance. Although the first reverse wave between 1924 and 1944 and the second wave of democratization during the mid-1940s and 1950s may still be distinguished, they are nevertheless not as convincing as seemed apparent in Figures 1 and 2. Moreover, there is no clear second reverse wave of democratization, which, according to Huntington, should have been apparent between 1957 and 1973. In fact, this period is better described as one of trendless fluctuation, in which there are waves of both authoritarianism and democracy.

CONCLUSION

This analysis has investigated the extent of the variation in transitions of political regimes that has occurred across different countries and over time. In studying the aggregated percentages of democratic regimes in the world over time, one can distinguish the three waves initially identified by Huntington. When the requirement of inclusive suffrage is imposed, however, the first wave is seen to begin much later, indicating that transitions to minimal democracy are a twentieth-century phe-

TABLE 1
REGIME TRANSITIONS PER WAVE

	# Transitions to Democracy	*# Transitions to Authoritarianism*	*# Transitions Outnumbered*[a]	*# Transitions Outnumbered,*[a] *Weighted by Countries*
First wave, 1893–1924	16	0	16	39
Reverse wave, 1924–44	3	7	–4	–9
Second wave, 1944–57	15	3	12	22
Fluctuation, 1957–76	16	18	–2	–3
Third wave, 1976–90	24	12	12	13
Explosive wave, 1990–94	34	4	30	31

SOURCES: For full details, see the appendix.
[a] Transitions outnumbered refers to transitions to democracy minus transitions to authoritarianism.

nomenon. Moreover, the peaks of the first and second wave of democratization also appear to be lower. Nor has this research confirmed Huntington's statement that there is a less than optimistic two-step-forward, one-step-backward pattern. Although the trend toward democratization may not be irreversible, the long-term trends in Figures 1 and 2 do clearly point to more long-term progress than Huntington has suggested.

This analysis also suggests that the focus on percentages of transitions may prove quite misleading. Since these percentages are susceptible to changes in the numbers of world states as well as in the numbers of transitions themselves, it is argued that a more accurate and meaningful impression of the waves of democratization can be achieved by plotting the real numbers involved in both transitions from authoritarian to minimal democracy and vice versa. These numbers may also be standardized to facilitate a genuine comparison over time. This new approach suggests that while the first wave of transitions to democracy still appears to be very striking, there is no longer any strong evidence of a second reverse wave, and the explosion of democratization in the period 1990–94, in which an impressive total of thirty-four authoritarian regimes effected a transition to minimal democracy, emerges with real force.

Why are these findings relevant? First, this article shows how im portant it is to have a clear conceptualization and classification of "dem ocratic regime." Scholars interested in historical processes of regim changes in order to identify waves of democratization should alway use a consistent distinction between democratic and authoritaria regimes. In addition, they should pay attention not only to the dimen sion of competition but also to the dimension of inclusiveness. This lat ter dimension is particularly important prior to the 1970s, when i discriminates powerfully between regimes. Since the 1970s, by contras almost all competitive regimes have been characterized by inclusive suf frage. By taking into account the dimension of inclusive suffrage, it ca be seen that the first wave of democratization starts later and the sec ond reverse wave is less distinctive.

Second, the findings in this research note emphasize that future re searchers should be more attuned to the implications of the changin number of states over time. Researchers who focus on *percentages* of de mocratic regimes among the states of the world assume a fixed underly ing geography; they typically think that the topology remains constan World history, however, has witnessed the emergence of many ne states, for example, as a consequence of the African decolonization i the 1960s and the independence of the former Soviet republics in 1991 By calculating the actual *number* of states (instead of percentages) tha made a transition from nondemocracy to democracy or vice versa, on arrives at a better insight into the real waves of democratization.

Third, these results indicate that future studies should be careful i comparing and explaining different waves of democratization. Hunt ington himself recommends such a comparison: "If the third wave o democratization slowed down or came to a halt, what factors may pro duce and characterize a third reverse wave? The experience of the firs and second reverse wave may be relevant."[33] Although it is reasonabl to expect that prior experiences suggest potential causes of future de velopments, future studies in which reverse waves are compared will b useless, because there are *no* clear reverse waves.

Finally, these findings are important not only because they suggest somewhat altered sequencing and a more accurate count of democrati zation waves but also because they imply real doubt about whether th wave metaphor is the most appropriate way to conceptualize the prob lem. Since reverse waves are not really apparent from these data, it ma be better to think in terms of "steps" toward democratization. There ar

[33] Huntington (fn. 1), 292.

certainly flows, but the ebbs are much less evident than had been averred. The way one counts influences how one thinks about the prospects for the continued expansion of democracy in the world. Many researchers simply expect a reverse wave in the near future and are waiting for it because they think that each wave is inevitably followed by a reverse wave. Are we on the edge of such a reversal? On the basis of the findings reported in this article, we know now that it is empirically possible for a wave of democratic expansion to be followed for some time not by a reverse wave but rather by an equilibrium in which the overall number of democracies in the world neither increases nor decreases significantly. It seems then that a period of trendless fluctuation is empirically more likely than a reverse wave.

APPENDIX: POLITICAL REGIME CHANGES, 1800–1994

A complete listing of regime changes is given in this appendix. Following each country is a list of all regime changes identified for that country. D indicates that a minimal democracy was established; regimes are considered as minimal democratic when they fulfill both the requirement of inclusiveness (measured by codings of the author) and the requirement of competition (measured by codings of Gurr's Polity III data set). A indicates that an authoritarian regime was established; regimes are considered as authoritarian when they do not fulfill the requirement of inclusiveness or the requirement of competition. I represents an interruption, interregnum, or transition period (measured by code –77, –66, –88 from Gurr's Polity III data set). Germany refers to Germany before 1948 and to West Germany after 1948; Czechoslovakia refers toCzechoslovakia before 1993 and to the Czech Republic after 1993.

Afghanistan 1800 A, 1992 I
Albania 1914 A, 1915 I, 1925 A, 1939 I, 1945 I, 1946 A, 1992 D
Algeria 1963 A
Angola 1975 A
Argentina 1825 A, 1830 I, 1835 A, 1852 I, 1852 A, 1946 I, 1948 A, 1955 I, 1957 A, 1973 D, 1976 A, 1983 D
Armenia 1991 D
Australia 1901 A, 1902 D
Austria 1800 A, 1805 I, 1806 A, 1809 I, 1810 A, 1818 I, 1920 D, 1933 I, 1934 A, 1938 I, 1945 I, 1946 D

Azerbaijan 1991 A
Baden (1818–71) A
Bahrain 1971 A
Bangladesh 1972 D, 1974 A, 1991 D
Bavaria (1800–1871) A
Belarus 1991 D
Belgium 1831 A, 1914 I, 1915 A, 1940 I, 1945 A, 1949 D
Benin 1960 A, 1963 I, 1965 A, 1990 I, 1991 D
Bhutan 1907 A
Bolivia 1826 A, 1841 I, 1842 A, 1871 I, 1873 A, 1952 I, 1956 A, 1982 D
Bosnia 1992 I
Botswana 1966 A, 1969 D,
Brazil 1824 A, 1930 I, 1934 A, 1945 I, 1946 A, 1964 I, 1965 A, 1985 D
Bulgaria 1879 A, 1913 I, 1915 A, 1942 I, 1944 I, 1947 A, 1990 D
Burkina Faso 1960 A, 1977 I, 1978 D, 1980 A
Burundi 1961 A, 1965 I, 1966 A
Cambodia 1950 A, 1953 I, 1955 A, 1970 I, 1972 A, 1975 I, 1976 A, 1991 I, 1993 D
Cameroon 1961 A
Canada 1867 A, 1921 D
Central African Republic 1962 A, 1993 D
Chad 1962 A, 1978 I, 1979 I, 1985 A
Chile 1818 A, 1924 I, 1925 A, 1955 D, 1973 A, 1990 D
China 1800 A, 1860 I, 1862 A, 1911 I, 1912 A, 1913 I, 1914 A, 1939 I, 1946 A
Colombia 1832 A, 1860 I, 1861 A, 1958 D
Comoros 1975 A, 1990 D, 1994 A
Congo 1961 D, 1963 A, 1991 I, 1992 D
Costa Rica 1838 A, 1919 I, 1920 A, 1949 D
Croatia (1830–1915) 1830 A, 1860 I, 1861 A, 1868 I, 1869 A
Croatia (1991–94) A
Cuba 1902 A, 1916 I, 1918 A, 1952 I, 1955 A, 1959 I, 1961 A
Czechoslovakia 1918 A, 1920 D, 1939 I, 1945 D, 1947 I, 1948 A, 1968 I, 1969 A, 1990 D
Democratic People's Republic of Korea 1948 A
Democratic Republic of Vietnam 1954 A
Denmark 1800 A, 1901 I, 1915 A, 1920 D, 1940 I, 1945 D
Dominican Republic 1844 A, 1861 I, 1865 A, 1914 I, 1925 A, 1930 I, 1932 A, 1961 I, 1962 A, 1963 I, 1964 A, 1978 D
Ecuador 1830 A, 1980 D

Egypt 1811 A, 1882 I, 1923 A, 1928 I, 1930 A, 1935 I, 1936 A, 1952 I, 1953 A
El Salvador 1841 A, 1855 I, 1858 A, 1948 I, 1950 A, 1979 I, 1984 D
Estonia 1917 A, 1919 I, 1920 D, 1933 I, 1936 A, 1941 I, 1988 I, 1991 A
Ethiopia 1855 A, 1936 I, 1942 A, 1991 I, 1994 A
Fiji 1970 D, 1987 A
Finland 1916 A, 1917 D, 1930 I, 1931 A, 1944 D
France 1800 A, 1851 I, 1852 A, 1860 I, 1863 A, 1870 I, 1877 A, 1944 I, 1946 D, 1958 A, 1969 D
Gabon 1961 A, 1990 I, 1991 A
Gambia 1965 A, 1966 D, 1994 A
Georgia 1991 A
German Democratic Republic (1949–89) A
Germany 1800 A, 1907 I, 1813 A, 1868 I, 1871 A, 1918 I, 1919 D, 1933 A, 1946 I, 1949 D
Ghana 1960 A, 1969 I, 1970 A, 1978 I, 1979 D, 1981 A, 1991 I, 1992 A
Gran Colombia 1821–30 A
Greece 1827 A, 1862 I, 1864 A, 1917 I, 1920 A, 1922 I, 1924 A, 1941 I, 1944 A, 1974 I, 1975 D
Guatemala 1839 A, 1871 I, 1873 A, 1879 I, 1880 A, 1985 I, 1986 D
Guinea 1958 A
Guinea-Bissau 1974 A, 1994 D
Guyana 1966 A, 1992 D
Haiti 1820 A, 1915 I, 1918 A, 1946 I, 1950 A, 1986 I, 1988 A, 1990 D, 1992 A, 1994 D
Honduras 1839 A, 1853 I, 1854 A, 1907 I, 1908 A, 1912 I, 1913 A, 1919 I, 1920 A, 1924 I, 1925 A, 1981 I, 1982 D
Hungary 1867 A, 1918 I, 1919 A, 1944 I, 1948 A, 1956 I, 1957 A, 1990 D
celand 1918 A, 1934 D
ndia 1950 A, 1952 D
ndonesia 1945 A
ran 1800 A, 1906 I, 1921 I, 1925 A, 1953 I, 1955 A
raq 1924 A
reland 1922 A, 1923 D
srael 1949 D
taly 1861 A, 1922 I, 1928 A, 1943 I, 1945 I, 1948 D
vory Coast 1960 A
amaica 1959 D
apan 1800 A, 1858 I, 1868 A, 1945 I, 1952 D

Jordan 1946 A
Kazakhstan 1991 A
Kenya 1965 A
Korea (1800–1910) A
Kuwait 1963 A, 1990 I, 1991 A
Kyrgyzstan 1991 D
Laos 1954 I, 1958 A, 1971 I, 1975 A
Latvia 1920 D, 1934 A, 1941 I, 1991 A
Lebanon 1944 A, 1971 D, 1976 I, 1990 A
Lesotho 1966 D, 1970 A, 1993 D
Liberia 1847 A, 1990 I
Libya 1951 A
Lithuania 1918 A, 1926 I, 1928 A, 1941 I, 1991 D
Luxembourg 1867 A, 1919 D, 1940 I, 1945 D
Macedonia 1991 D
Madagascar 1961 A, 1992 I, 1993 D
Malawi 1965 A, 1994 D
Malaysia 1957 D, 1969 A, 1971 D
Mali 1960 A, 1991 I, 1992 D
Mauritania 1961 A
Mauritius 1969 D
Mexico 1822 A, 1834 I, 1835 A, 1846 I, 1848 A, 1863 I, 1864 A, 18 I, 1880 A, 1911 I, 1917 A
Modena (1815–60) A
Moldavia 1991 A
Mongolia 1924 A, 1993 D
Morocco 1800 A, 1912 I, 1956 A
Mozambique 1976 A, 1994 D
Myanmar (Burma) 1948 D, 1962 A
Namibia 1990 D
Nepal 1800 A, 1958 I, 1959 A, 1991 D
Netherlands 1815 A, 1922 D, 1940 I, 1945 D
New Zealand 1857 A, 1893 D
Nicaragua 1838 A, 1926 I, 1927 A, 1979 I, 1981 A, 1990 D
Niger 1959 A, 1991 I, 1993 D
Nigeria 1960 A, 1978 I, 1979 D, 1984 A
Norway 1814 A, 1921 D, 1940 I, 1945 D
Oman 1800 A, 1957 I, 1959 A
Orange Free State (1854–1902) A
Pakistan 1947 A, 1950 D, 1958 A, 1993 D
Panama 1903 A, 1990 D

Papal States (1815–70) A
Papua New Guinea 1976 A, 1977 D
Paraguay 1816 A, 1868 I, 1870 A, 1989 D
Parma (1815–60) A
Peru 1821 A, 1825 I, 1828 A, 1881 I, 1882 A, 1919 I, 1920 A, 1930 I, 1933 A, 1978 I, 1980 D, 1992 A
Philippines 1935 A, 1941 I, 1945 A, 1950 D, 1969 A, 1986 I, 1987 D
Poland 1918 A, 1922 D, 1926 A, 1939 I, 1948 A, 1991 D
Portugal 1800 A, 1801 I, 1802 A, 1807 I, 1820 I, 1823 A, 1833 I, 1834 A, 1910 I, 1911 A, 1926 I, 1930 A, 1974 I, 1976 D
Republic of Korea 1948 A, 1960 D, 1961 A, 1988 D
Republic of Vietnam (1955–75) A
Romania 1859 A, 1916 I, 1917 A, 1940 I, 1941 A, 1944 I, 1948 A, 1989 I, 1990 D
Russia 1800 A, 1812 I, 1813 A, 1905 I, 1906 A, 1991 D
Rwanda 1960 A, 1994 I
Sardinia (1815–60) A
Saudi Arabia 1926 A
Saxony (1806–71) 1806 A, 1813 I, 1814 A, 1848 I, 1849 A
Senegal 1960 A, 1962 I, 1964 A
Sierra Leone 1961 D, 1967 A
Singapore 1959 D, 1963 I, 1965 A
Slovakia 1993 D
Slovenia 1991 D
Somalia 1960 A, 1964 D, 1969 A, 1991 I
South Africa 1856 A, 1902 I, 1908 I, 1910 A, 1993 I, 1994 D
Spain 1801 A, 1808 I, 1814 A, 1836 I, 1837 A, 1869 I, 1871 A, 1874 I, 1876 A, 1930 I, 1931 D, 1939 A, 1975 I, 1978 D
Sri Lanka 1948 D, 1977 I, 1978 D, 1982 A, 1994 D
Sudan 1954 A, 1964 I, 1965 D, 1969 I, 1971 A, 1985 I, 1986 D, 1989 A
Swaziland 1968 A
Sweden 1800 A, 1809 I, 1812 A, 1907 I, 1917 A, 1921 D
Switzerland 1848 A, 1971 D
Syria 1944 D, 1949 A, 1954 D, 1958 I, 1961 A
Taiwan 1949 A, 1991 D
Tajikistan 1991 A
Tanzania 1963 A
Thailand 1800 A, 1932 I, 1935 A, 1941 I, 1942 A, 1968 I, 1969 A, 1973 I, 1975 D, 1977 I, 1978 A, 1988 D
Togo 1961 A, 1991 I, 1993 A
Trinidad 1962 D

Tunisia 1959 A
Turkey 1800 A, 1919 I, 1922 I, 1923 A, 1946 D, 1953 A, 1960 I, 1961 D, 1971 A, 1973 D, 1980 A, 1984 D
Turkmenistan 1991 A
Tuscany (1815–60) A
Two Sicilies (1816–60) A
Uganda 1962 D, 1966 I, 1967 A, 1979 I, 1981 D, 1985 I, 1986 A
Ukraine 1991 D
United Arabic Emirates 1971 A
United Kingdom 1800 A, 1928 D
United Provinces of Central America (1824–38) A
United States of America 1800 A,1920 D
Uruguay 1830 A, 1952 D, 1972 I, 1973 A, 1985 D
Uzbekistan 1991 A
Venezuela 1830 A, 1963 D
Württemburg (1800–1871) A
Yemen 1990 I, 1994 A
Yemen Arabic Republic (1918–89) 1918 A, 1946 I, 1948 A
Yemen People's Republic (1967–89) A
Yugoslavia 1921 A, 1937 I, 1939 A, 1941 I, 1944 I, 1946 A, 1951 I, 1953 A
Zaire 1960 I, 1965 A, 1993 I
Zambia 1964 A, 1991 D
Zimbabwe 1923 A, 1979 I, 1980 D, 1983 A

Review Article

The Study of Money

By JONATHAN KIRSHNER*

Benjamin J. Cohen. *The Geography of Money.* Ithaca, N.Y.: Cornell University Press, 1998, 229 pp.

Susan Strange. *Mad Money: When Markets Outgrow Governments.* Ann Arbor: University of Michigan Press, 1998, 212 pp.

MONETARY phenomena now define the contours of the contemporary global economy. This is new, and it has transformed the study of international political economy (IPE). By contrast, for most of the past two centuries, money, as John Stuart Mill argued (and like a good baseball umpire), was only noticeable when it failed. This was certainly the case during the dramatic expansion of the international economy following the Second World War, and it was reflected in the concurrent development of the subdiscipline of IPE. The real economy —trade, especially liberalization, integration, and strategy—dominated the agenda, while monetary politics tended to pop up when its more severe pathologies—sterling crises, the collapse of Bretton Woods, the dollar cycle—threatened to disrupt the smooth functioning of world trade.

At the dawn of the twenty-first century, however, these roles have reversed. Money leads and the real economy must follow. From an economic perspective, this has an Alice-in-Wonderland quality to it: money, once the handmaiden of the modern economy, has become its mistress. For students of IPE, this presents a fundamental and profound question: in what way are politics affected by this change?

Two excellent books, *The Geography of Money*, by Benjamin Cohen, and *Mad Money*, by Susan Strange, will frame, support, and serve as the point of departure for scholars addressing this vital question. The volumes, by two of the most accomplished and distinguished experts on

* I thank Rawi Abdelal, Eric Helleiner, Charles Kindleberger, Karl Mueller, two anonymous referees, and the editors at *World Politics* for helpful comments and suggestions.

World Politics 52 (April 2000), 407–36

the IPE of money, also establish landmarks around which competing schools of thought will organize. Cohen, the thoughtful liberal, trusts markets but understands that monetary phenomena have inescapably political consequences. Strange, the skeptical Keynesian, recognizes the indispensable power and efficiency of markets but sees a monetary order reeling from political conflict and ubiquitous market failure.

From these distinct perspectives, each author provides a new roadmap for navigating the international political space transformed by the ascendance of money. They also remind us why so many of the great economists, including Ricardo, Marx, Marshall, Keynes, Schumpeter, Myrdal, and Hayek, dedicated much of their efforts to the study of monetary phenomena. Deceptively theoretical debates about money—its definition, management, and idiosyncrasies—are typically rooted in fundamental conceptions of politics and society.

Ultimately, however, and perhaps necessarily, these books raise more questions than they answer. But they do suggest a direction for the most promising avenues of investigation—toward the study of the unique interconnections between the ideas, material interests, and institutions associated with the management of money. Those relationships are profoundly consequential for politics and demand the renewed attention of contemporary scholars of international relations and political economy. Cohen, and more explicitly Strange, clearly acknowledge the importance of ideas in monetary affairs. Cohen reminds the reader that money derives its very value from shared social conventions, indeed, "has no meaning at all except with reference to the mutual confidence that makes its use possible" (p. 11). Strange is even more pointed, asserting that the basic differences between competing economic theories are "not technical but political" (p. 190).

But while the role of ideas is an important implicit foundation of these volumes, neither takes on the issue in a systematic way, largely because each has a dominant theme. These themes, reflected in their respective titles, provide the nominal touchstone for their analyses. This article, therefore, proceeds in two parts. The first part considers *Geography* and *Mad Money* on their own terms, concentrating on the themes stressed by the authors. The second part of the article then explores what is left implicit by these volumes—the need for a renewed attention to ideology and politics in explaining monetary phenomena.

This latter part of the article first shows how ideas can play a unique role in shaping monetary phenomena. It then argues that this matters because those ideas can mask distributional conflicts. Fundamentally political struggles about money, that is, are routinely cloaked in eco-

nomic terms, often throwing students of politics off the scent. This discussion then concludes with two illustrations of macroeconomic policies—the pursuit of low inflation and the deregulation of capital flows—that have been shaped by ideas that are of greater political than economic consequence.

THE NEW GEOGRAPHY

As Cohen notes in his preface, he traces his interest in the geography of money to his first professional publication.[1] Even more broadly, *Geography* can be seen as the most recent and arguably the culmination of a series of important books by Cohen, including *The Future of Sterling as an International Currency*, *Organizing the World's Money*, and *In Whose Interest? International Banking and American Foreign Policy*. Each of these books negotiated the terrain between economics and politics as it also explored crucial issues relating to international money and finance. Thus *Geography* breaks new ground while also tying together arguments and analytical themes that have characterized Cohen's writing throughout his career.

Cohen's premise is straightforward. While the global financial revolution is often characterized as undermining state authority in general and national control over money in particular, "the geography of money is far more complex than we generally assume" (p. 1). Cohen quickly establishes a clear thematic foil: the myth of one nation, one money, that is, the idea that within their own borders, states are their own monetary masters. This has never been true, he argues, and is less and decreasingly an approximation of reality now than at any other time in recent history. Cross-border currency competition is a fact of life and is continuously eroding the power of governments. And this matters, because "international relations, political as well as economic, are being dramatically reshaped by the increasing interpenetration of national monetary spaces" (p. 3).

To comprehend the meaning of these changes, Cohen argues, we must understand the geography of money. The most important factor driving change in that geography is the shift in power from states to markets. This shift, however, is not a transformation from state supremacy to decentralized globalization. Rather, it is a change in the na-

[1] Benjamin J. Cohen, "The Euro-Dollar, the Common Market, and Currency Unification," *Journal of Finance* 17 (December 1963), esp. 613–15. In this paper Cohen argued, among other things, that the existence of the euro-dollar market was an important reason why the members of the European Economic Community had not developed a common currency.

ture of state power. While at one time states could be considered the monopoly producers of currency within their borders, they have now been reduced to (still powerful) oligopolists. They retain the capacity to shape the market but also find themselves constrained by competitors.

With his foil, the one nation, one money myth, and his question, the political and economic consequences of increased interpenetration of international monetary spaces, Cohen proceeds with his argument systematically and effectively. First, looking back, *Geography* explores the old-fashioned world of territorial money from a historical perspective. While there was always sovereign coinage, Cohen notes, the conceptualization of money in territorial terms did not come about until the nineteenth century. Before then foreign coins circulated with little concern for national boundaries. But leaders overseeing the consolidation of the nation-state in the nineteenth century used money as in instrument of state building,[2] and the establishment of national money as "legal tender" drove out foreign (currency) competition.

In retrospect, then, the century from 1870 to 1970 was an atypical interlude in which states enjoyed a near monopoly in currency affairs. This monopoly gave states power vis-à-vis other societal actors in four ways. First, it provided a powerful political symbol, which is one reason why newly independent states today continue to establish distinct currencies despite their reduced economic significance. Second is the opportunity for seigniorage, which can be an important source of revenue for states and, importantly, the opportunity for which is inversely related to the ability of subjects to substitute into another currency. States that control their money can also practice macroeconomic management by manipulating the money supply and the exchange rate—important policy tools, especially in the short run. Finally, by issuing their own currency, states gain power in a negative sense: they avoid dependence on some other source. For these reasons states continue to support the Westphalian conception of one nation, one money, even though today "it is no more than a myth" (p. 46).

Not only is the Westphalian conception of money a myth, but Cohen illustrates that it was an oversimplification even during its heyday. While states had more control over their currencies from 1870 to 1970, they nevertheless routinely chose to abdicate some of that sovereignty, in two ways—either by subordinating or by sharing their monetary sovereignty. States routinely subordinated their sovereignty,

[2] On this point, see Eric Helleiner, "Historicizing Territorial Currencies: Money Space and the Nation-State in North America," *Political Geography* 18 (March 1999).

choosing policies that range from wholesale abandonment through the adoption of a currency board to the more modest participation in a monetary bloc. States have also chosen to share their monetary sovereignty by forming currency unions, trading some of the benefits of a monopoly over money in exchange for improved autonomy vis-à-vis the outside world, making more efficient use of money (in its classic economic roles) within the monetary union, and forming a broader political symbol. While Cohen's account of these exceptions to the Westphalian myth are interesting in their own right, they are also of increased practical significance, as all states must now weigh their options in the face of reduced monetary sovereignty.

Cohen next offers a rich empirical examination that provides the bridge between the first half of the book, which sets out the argument, and the following three chapters, which explore its implications. Looking at the contemporary global economy, he attempts to assess the extent of currency internationalization and currency substitution, providing support for the assertion that one nation, one money is a myth. "The picture that emerges from this survey is one of intense competition as well as distinct hierarchy among the world's many currencies" (p. 93). Cohen then considers this hierarchy and introduces a new typology, the "Currency Pyramid," to classify the role of different currencies and to help visualize a new landscape in which most currencies face rivals but few are used internationally. Featuring seven categories, ranging from top and patrician currencies all the way down to quasi and pseudo currencies, this typology can be seen as an effort to update Strange's classic descriptions of thirty years ago.[3]

Having established his empirical claim and sketched the new geography of money, Cohen offers three chapters that consider the political implications of the new world: these are the core of what will be the book's most lasting contributions. The first, "A New Structure of Power," revisits the four elements of power—symbolism, seigniorage, management, and insulation—that states derived from issuing currency during the Westphalian era. He asks, "What happens to structures of power when currencies are no longer territorial?" (p. 119). Two themes dictate what is new for each of the four elements. First, the changes will be defined by the new relationships between state and market, not between states. States that are sensitive to market sentiment will be able to retain aspects of the four elements. Second, while all states will find themselves increasingly constrained by the market, those at the bottom

[3] Susan Strange, "The Politics of International Currencies," *World Politics* 23 (January 1971).

[illegible] fewer options than those at the top. Thus, [illegible] the political symbolism to be derived from [illegible] own currencies, so long as they manage those issues in an appropriate fashion. Indeed good management, yielding a "strong currency," can "enhance a government's reputation" (p. 121). Similarly, the prospects for macroeconomic management will depend increasingly on "how official policies interact with market preferences" (p. 125). With regard to seigniorage and insulation, the emphasis is on asymmetry: those at the top of the pyramid will find their positions enhanced (so long as they do not drift too far from market preferences), while those at the bottom will lose out. The biggest changes, however, as Cohen reminds us, are not about the differences between states but rather concern the fact that governments in general will have less power in relation to other social actors than they had in the distant past. The big winners politically are those private social actors who can choose their preferred transaction network.

Chapter 7, "Governance Transformed," considers this very question: the role of the state, especially vis-à-vis private actors, in the new monetary landscape. Here Cohen attempts to seize the middle ground. While states are weaker than they once were, they are surely not powerless. His main point is that "the shift in the structure of power generated by cross-border currency competition has not so much diminished as *transformed* the role of the state in money's newly deterritorialized geography. Governance is now uneasily shared between the public and private sectors" (p. 131). Privileged elements in the private sector now have an easier exit option, and governments are under pressure not to provoke such exit. But Cohen argues that this is only part of the picture, because it restricts itself to the demand side of the equation. Governments, as the providers of currency, represent the vital supply side, a role that has not been eliminated by the transformation. Rather, governments have been transformed from monopolists to oligopolists. "States, once largely supreme in their own territories, have now become something like competing firms in an oligopolistic industry" (p. 138). And as oligopolists, governments can act to preserve and promote the market share for their products. Governance, once provided by states, is now in the hands of the market forces that shape the strategies states must pursue in the context of their oligopolistic competition.

Cohen the liberal economist has confidence in the efficiency of market governance, but Cohen the political scientist raises grave doubts about two aspects of the new geography of money. First, there is the problem of equity: some actors will get much more voice in the new

in particular, the wealthiest segments of society. Second, there is of accountability: however much politicians may be objects of at the very least they could be held accountable for their actions. et forces, by contrast, "are neither elected nor politically account- (p. 148).

..ese crucial questions of equity and accountability are not, however, the focus of the final chapter, "Can Public Policy Cope." For this ason, of the three grand chapters that close this book, this one is disppointing in that it does not fulfill the agenda raised by what came before it. Cohen asserts, accurately, that there are "no easy solutions." The bottom line ultimately is that "governments must learn how to *manage* their oligopolistic rivalry, not make futile attempts to evade or suppress it" (p. 151). Cohen then shifts gears and provides an overview of the new monetary landscape, emphasizing that neither a reassertion of state power nor a complete denationalization of money is likely to occur. Rather, exploring each level of the currency pyramid, including the euro, the yen, and the bottom dwellers, he concludes that the world is likely to continue to look pretty much the way it does now: with the gentle erosion of the still dominant dollar tempered increasingly not by competition from the euro or the yen but by the power of markets.

CAN THE MIDDLE HOLD?

Cohen's treatment of money, though politically sensitive and sophisticated,[4] nevertheless represents the economistic pillar of the IPE mainstream. For this branch of the subfield, and especially with regard to the study of money, the use of market metaphors and the assumption of market efficiency must be carefully scrutinized.

The use of economic metaphor is one of Cohen's key contributions. Between Westphalia and globalization, he argues, there is a middle ground of oligopoly—an attractive and novel conceptualization that would appear to strike a clear middle ground between two opposing schools. Issuers of currency can easily be cast as traditional oligopolists.

[4] Cohen notes, for example, that the extent to which the yen is used as an international currency is essentially a policy choice and that both the Europeans and the Japanese manage their currencies with an eye toward avoiding confrontation with the United States (pp. 161, 164). He also assesses monetary geography in its most basic form, sustaining monetary alliances, and finds that economic theory, such as the theory of optimal currency areas, has little explanatory power. Rather, political factors such as power and ideology explain which monetary unions succeed and which fail (pp. 82, 84–85, 87, 91). See also the cases discussed in Cohen, "Beyond EMU: The Problem of Sustainability," *Economics and Politics* 5 (July 1993). He also traces the drive for EMU to political motives, but it should be noted that he reviewed the convergence criteria without a discussion of their possible distributional implications (*Geography of Money*, 74, 77).

Operating as they do in a world of interdependence and uncertainty they must focus on promoting their brand name and assuring the market share for their product.

But are the dollar, yen, and euro like Boeing, Coke, and Kodak? In at least one crucial way, the answer is no. It is not the obvious difference between public and private that undermines the analogy but rather the difference between the real and monetary sides of the economy. Boeing, for example, has orders and contracts that will take years to fulfill. It also has physical plants and equipment, technical know-how, and a small number of identifiable competitors. Entry and exit for producers and consumers in the aerospace industry, while certainly possible, is a relatively slow process. Thus, if Boeing has a bad quarter or even a few bad quarters, it may lose some of its market share. But it will lose only so much. There will be no "flight from Boeing" (as it were), even if it faced a great crisis, as might occur if one of its planes were found to have a design flaw.

But currencies are different. Ultimately, they are worth, well, what people think they are worth. And if confidence wavers, people can get rid of that asset and almost immediately acquire any of dozens of alternative assets. Further, those very actions may be self-fulfilling—creating expectations of further flight and a "rational" erosion of confidence. Cohen himself is quite explicit about the ephemeral nature of a currency's attractiveness (p. 97, and also pp. 10–13)—it is attractive as long as it conforms to those policies that the market finds attractive. As Cohen notes, reputation, which comes at "considerable cost and effort," is crucial (p. 141).

But this raises a fundamental point. While the international use of currency may look like an oligopoly (concentrated use in a few brands), this does not mean that the producers enjoy oligopolistic power. What distinguishes oligopolists from other actors in a market economy is that they are not mere "price takers." Rather, they can affect the market and the price for their good—in a much more limited fashion than monopolists, to be sure, but still to an extent that distinguishes them sharply from the textbook firm. But in currency affairs exit is so easy and the extent to which states are adhering to market preferences is so transparent that issuing states in fact may have very little leeway to act as oligopolists.

Cohen's own arguments make this clear. Efforts to protect a currency's reputation will be successful if they "make significant concessions to market sentiment" (p. 122). This is true even for states at the very top of the currency pyramid. Cohen had argued that such states

could still reap the political benefits afforded by national money, such as seigniorage and the ability to practice macroeconomic management; in both cases such autonomy is circumscribed by the need to "retain competitive superiority in the marketplace" and by "how official policies interact with market preferences" (pp. 123, 135). Thus Cohen's oligopolists may face the constraints of perfect competition.[5]

This is critical for Cohen's analysis, because if the wiggle room of the oligopolist is eliminated, then we must confront the possibility that we are de facto in the purely globalized world where states have no power and markets rule. While Cohen raises political concerns about markets (equity and accountability, as mentioned above), he still holds to a fundamental faith that unfettered market forces will yield optimal outcomes. Smaller states, for example, may achieve "a much healthier economic performance" if they "in effect submit their national sovereignty, at least in part, to the strict discipline of the marketplace" (p. 126).

But what if laissez-faire is suboptimal when it comes to international money? This, one of the fundamental divergences between Cohen and Strange, remains an unresolved question; it will be addressed at length below. Cohen does offer at least one good deductive reason to suspect that unregulated money is not a good idea. He notes, for example, that one technique of bolstering a currency's reputation is to raise interest rates. But in a world of oligopolistic competition, where reputation is relative, this may lead to competitive interest-rate increases—for political reasons, rather than for economic ones.[6] The result—suboptimally high interest rates having no economic justification—would slow global economic activity.

The prospect of flaws in the functioning of the international macroeconomy is the weak link of Cohen's analysis and, by extension, of the study of money from this perspective in the IPE field as a whole. Cohen explicitly embraces the market, and quite pointedly so, explaining that as state power erodes, we are left not with anarchy but with governance by the market. Indeed, responding directly to the assertion by Strange that states have been replaced by anarchy, Cohen answers that "nothing

[5] Cohen recognizes this, explaining that autonomy is likely to erode even for the most powerful states and noting that public policy will have to conform increasingly with what markets want (pp. 130, 133). His faith in market efficiency masks the tension between these observations and his oligopoly analogy.

[6] For an illustration of the potentially disastrous consequences of interest-rate competition, derived in part from a concern with market reputation, see Barry Eichengreen, *Golden Fetters: The Gold Standard and the Great Depression* (Oxford: Oxford University Press, 1992). See also Jonathan Kirshner, "Disinflation, Structural Change, and Distribution," *Review of Radical Political Economics* 30 (March 1998).

could be more mistaken" (p. 142). The market, like the state, according to Cohen, is a socially constructed system of authority. Thus, governments have been replaced "not by anarchy but by the invisible hand of competition" (p. 146).[7]

For scholars raised on a steady diet of IR theory, this will be the one awkward moment of the book. For while markets may be socially constructed, once formed they do create an anarchic environment—that is, an environment that lacks central authority. In its purest form, while "the market" is the result of the sum of actions taken by individual actors, those actions combine to create constraints on each one that no single actor can affect. It is this very market analogy and the appeal to anarchy that was the foundation of Kenneth Waltz's defense of systemic theorizing in international politics.[8] Additionally, the implicit appeal to Adam Smith in this argument is also problematic, because Smith did not believe that all monetary matters should be left to the market. He supported, for example, regulations to control interest rates.[9]

Thus, despite Cohen's thoughtful ruminations on the nature of "governance," it is hard to avoid the conclusion that as state power erodes, we are left increasingly to anarchic (though not necessarily chaotic) market forces. As noted, Cohen sees political consequences flowing from this change; such an arrangement, he states, faces a "crisis of legitimacy" (p. 147). But ultimately, a basic faith in the market underpins *Geography* and that branch of the subfield which it ably represents.

A Mad, Mad World

Susan Strange's *Mad Money*, like Cohen's volume, is a major statement that integrates the accumulated knowledge of a foremost scholar of the IPE of money. But beyond that, the similarities end. Even as Cohen refines themes and applies them in different contexts, *Geography* nonetheless reflects more continuity than change. *Mad Money*, by contrast, completes a transition in Strange's work. Her early efforts, such

[7] In a parallel argument to his oligopoly metaphor, Cohen argues, with an appeal to Alfred Marshall, that governance that emerges from the market reflects the outcome of both demand and supply-side forces. But, following Marshall, it is the shape of the curves that is determinant. If the demand curve takes the shape of a flat line (is infinitely elastic), then supply has no voice in determining the equilibrium price.

[8] Waltz, *Theory of International Politics* (New York: Random House, 1979). Interestingly, an outstanding application of systemic theorizing that illustrates the consequences of anarchy can be found in Benjamin Cohen, *The Question of Imperialism* (New York: Basic Books, 1973).

[9] Smith, *The Wealth of Nations* (1776; Chicago: University of Chicago Press, 1976), 379–80.

as the seminal *Sterling and British Policy*, as well as *International Monetary Relations, 1959–1971*, were market sensitive but statecentric. Later, in books like *Casino Capitalism* and *Rival States, Rival Firms*, market forces unleashed by states became the principal constraint upon individual governments and posed the most basic challenges to them. Finally, at the close of her career, in *The Retreat of the State* and with the publication of *Mad Money*, those market forces have triumphed.

Mad Money also differs from *Geography*, as noted above, in its much deeper skepticism of the consequences of those unleashed forces. Strange's perspective, not surprisingly, is also more overtly political: actors often have sharp, conflicting motivations, and political variables figure prominently in the foreground. The books are stylistically distinct as well. While Cohen constructs a careful, linear argument, Strange's book lays out its principal theme and then embarks on a thematic survey of the global terrain.

The title itself captures her argument. Why *mad* money? Because the definition of mad is "wildly foolish," and this is the proper way to characterize the decision "to let the financial markets run so far ahead, so far beyond the control of state and international authorities" (p. 1). Financial markets themselves have also turned out to be "mad": erratic, unpredictable, and dangerous. Strange observes pathological behavior, and her first two chapters thus diagnose the problem. Increases in the size of financial markets and the speed of financial transactions, along with the decrease of national and international supervision and regulation of banking and financial services, have resulted in a global financial system even more fragile and volatile than the one she described in *Casino Capitalism*. Strange then traces the series of decisions (and non-decisions) made by states that allowed these changes to occur. In her assessment, the fact that markets have outgrown governments is "*the* prime issue of international politics and economics" (p. 18, emphasis in original).

Strange's next two chapters focus on the political foundations of the international financial system. First is the crucial bilateral relationship between the United States and Japan: she sees reduced likelihood that the needed political cooperation between the world's two largest economies will take place. This discussion includes, in distinct sections, the most dated and the most farsighted analysis of the volume. The sources of tension are mainly the usual suspects from the late 1980s, when the end of the cold war, the surging Japanese economy, and enormous American budget deficits fostered an atmosphere of growing sus-

picion between the two states.[10] But on the heels of this review comes a broader discussion of the importance of international cooperation and its political underpinnings. For when crises occur, as they inevitably will, swift intervention is needed on the part of governments, not bankers. Central banks are ill-suited to the task, as they "have an innate tendency to prefer deflation to inflation, and to judge monetary stability more important than keeping people in jobs and troubled businesses from closing down" (p. 58). Here Strange explicitly invokes Keynes, praising his emphasis on the real side of the economy, in contrast with the deflationists of the interwar years.[11]

The second political foundation considered by Strange is the relationship between France and Germany. Politics—"the domestic politics of France and Germany and the impact of both on the relations between them" (p. 60)—will ultimately decide the fate of the euro. While international politics—the end of the cold war liberating German foreign policy from its dependence on America—has allowed the euro to come about, politics will also assure that it will be a continuing source of instability. On the road to unification the uncertainty that necessarily accompanies politics will result in recurrent bouts of speculative pressure. Afterward, because the European Union's political structure is not up to the task of managing the political conflicts that will emerge over the management of a common currency, the great European monetary project will struggle at first and then most likely fail.

From this perspective—shaped by the rise of mad money and the decline of international political cooperation—Strange moves on in the middle of the book to survey a number of disparate issues, such as stock market speculation, debt crises, and transnational organized crime. These issues are unified, however, by the consistent analytical perspective of "global Keynesianism." Strange never declares this, but her loyalties are unambiguous. Thus in one instance she characterizes a writer as a global Keynesian and then later notes that his position "echoes [her] own argument" (p. 14). She also goes out of her way repeatedly to praise Keynes's work and track record, and she champions policy prescriptions that are explicitly and implicitly associated with him.[12]

[10] It should be noted that Strange does touch on some more recent developments, such as the increased concentration of savings in Japan's postal savings system and the shift in the composition of Japan's U.S. holdings from long-term securities to short-term Treasury bills (p. 50).

[11] Strange also argues that protectionism and trade blocs are normally the result, not the cause, of economic distress, which can usually be traced to the financial system. As such, she contends, relatively too much attention has been given to the politics of trade (p. 58, see also pp. 87, 186).

[12] See, e.g., pp. 16, 88, 90, 92, 120, 183. These themes were repeated in one of Strange's last works, where she addressed the "most amazing omission" in contemporary IPE scholarship, the work of

One of the principal problems of the contemporary global economy, Strange argues, is the pressing need to apply Keynesian policies at a time when it is exceptionally difficult to implement them. This is because Keynes's "great work," *The General Theory*, provided a template for the management of national economies that are, if not closed, at least relatively insulated. Such a world no longer exists, however, due in large measure to the (ill-advised) dismantling of the capital controls that most states deployed during the Bretton Woods era.[13] (Recall that Keynes, one of the principal architects of the Bretton Woods system, wrote: "Nothing is more certain than that the movement of capital funds must be regulated.")[14]

Thus, by the mid-1980s "the logic of the *General Theory* still held. But henceforth if countercyclical intervention were to be effective, it would have to be global and collective, not national" (p. 91). Not only is the cooperation needed for such collective action less likely, however, but few powerful states or institutions have the inclination to support such policies anyway. Big countries like the United States see too few short-run interests at stake to justify the effort, while the IMF's philosophy is "monetarist rather than Keynesian, anti-inflationary, opposed to fiscal deficits, state ownership, and welfare subsidies" (p. 93). Because of this, the world has collectively botched the transformation of the former Soviet bloc and the management of the Mexican and Asian financial crises, at great real cost and with lasting and substantive consequence.[15] Relatedly, the orthodox treatment of financial crises in general is inadequate, since it focuses solely on the domestic economies of the states involved and not on the crucial international mechanisms through which financial panics can be fomented, exacerbated, transmitted.[16]

Keynes. See Susan Strange, "Finance in Politics: An Epilogue to *Mad Money*" (Draft paper prepared for *Power and Order: Change in World Politics—A Festschrift in Honor of Robert G. Gilpin*, Princeton University, October 2–3, 1998), 7 and the discussion on 8–9. See also Jonathan Kirshner, "Keynes, Capital Mobility, and the Crisis of Embedded Liberalism," *Review of International Political Economy* 6 (Autumn 1999).

[13] Obviously, the folly of capital deregulation is one of the main themes of Strange's book. See for example pp. 16–17; also idem (fn. 12), 6, 17–18.

[14] John Maynard Keynes, "Letter to Roy Harrod," April 19, 1942, reprinted in Donald Moggridge and Elizabeth Johnson, eds., *The Collected Writings of John Maynard Keynes* (London: Macmillan, 1971–89), 25:149.

[15] See pp. 103–6, 109–11, 118–20, and note especially Strange's distinction between nominal recovery, "the restoration of confidence in financial markets" (p. 105), and the performance of the real economy.

[16] See pp. 106, 112. Strange's perspective on the tendency for international financial instability and the need for international governance is similar to Kindleberger's, whom she also cites repeatedly and approvingly. See pp. 15, 55–56, 86; also Charles Kindleberger, *The World in Depression*, rev. and expanded ed. (Berkeley: University of California Press, 1986); and especially idem, *Manias, Panics, and Crashes*, 3d ed. (New York: Wiley, 1996).

Finally, neither of the two potential "international guardians" appears poised to save the day. Indeed, the Bank for International Settlements, while still a vital forum for communication between centra banks, has "abandoned, in effect, the whole idea of agreed commor standards of banking supervision" (p. 161). Nor is the IMF, even disregarding the criticisms raised elsewhere in the volume, up to the task On the contrary, as it faces greater problems, ever larger bailouts, and importantly, declining legitimacy, Strange finds a "chilling analogy witl the U.S. experience in Vietnam," with the fund "biting off more than i can chew" (p. 164).

In looking out the window, Strange sees much the same world tha Cohen describes. At century's end "the nation state is not up to the jol of managing mad international money" (p. 190) and has been superseded by an increased "reliance on the discipline of market forces" (p 176). And like Cohen, Strange argues that this will empower some a the expense of others: in a world dominated by vast international credi markets large firms and large states will do better than small firms an(small states.

But for Strange, the problems run much deeper. Unlike Cohen sh(does not view laissez-faire in money and finance as economically efficient. Its pathologies run both hot and cold, and Strange envisions tw(potential paths by which mad money could lead to a global economi(crisis. First by fire: irrationality, rooted in greed and fear (the "emotion most prevalent in the day to day behavior of the international financia system today" [p. 139]) and perhaps goaded by political instability ii Russia or China unleashes a financial panic that a sluggish political response fails to contain. Or second by ice: an exogenous contraction o credit, exacerbated by "the deflationary bias of monetarist economi(strategies . . . land the world production and trade structures in a viciou circle of falling demand" (p. 185) that, again, is inadequately contained Either way, mad money will come to no good. The only silver lining i that when a major crisis does occur, it will probably provide the necessary impetus for fundamental reform.[17]

THE ROAD FROM HERE

The great strengths of *Mad Money* lie in its diagnosis of the problen and the way that its infusion of politics animates each discussion. Bu Strange does not advance any specific theoretical apparatus and is cau-

[17]See, e.g., pp. 168, 190.

tious about making predictions. She is also quite realistic—that is, pessimistic—about the political prospects for "global Keynesianism" and offers few details as to how it would be managed in practice.[18] As a result, to an even greater extent than *Geography*, the book raises more questions than it answers.

But if Strange and Cohen do not provide the answers, they do offer guidance for those who would follow in their footsteps. While both *Geography* and *Mad Money* suggest dozens of avenues for investigation, there is one recurrent theme that points to a clear research trajectory. The implicit and to a large extent unintentional fulcrum of these books is the theme of a distinct relationship between ideas, institutions, and interests in the monetary sphere. A closer look at ideas about money leads ultimately to the study of conflict over money and from there more broadly to the politics of money. The following discussion, therefore, unfolds in four linked and interrelated stages. First, the profound significance of ideas in monetary affairs is considered, both for authors' perspectives and for the discipline as a whole. Crucially, ideas about money can have an independent effect not only on which policies will be chosen but also upon which types of policies will be successful. The discussion then turns to why this matters: even if monetary phenomena are neutral (with no long-run influence on aggregate economic performance), they can and typically do have sharp distributional effects. Thus many economic policies that are justified principally on the strength of their theoretical soundness (and superiority to other options) may be of predominantly political origin and effect. Two examples—inflation and capital mobility—are then introduced to illustrate this point. In each case, ideas have played a crucial role in shaping and sustaining one policy option, even though the evidence suggests that the economic distinctions between competing policies are relatively small and ambiguous, while the political consequences appear sharp and substantial.

IDEAS: MORE THAN YOU THINK

There is an important and growing literature on the importance of ideas in international relations, as well as in IPE, including some works that have addressed how ideas have shaped decisions made about

[18] One exception is her repeated plea for the closure of tax havens, one measure at least that should be feasible. See pp. 131, 188–89. On Strange's caution about efforts to make specific predictions in IR, see pp. 19–20, 179.

mains an essential analytical device for students of the international economy and of IPE. One of the understood "laws" of economic theory derived from Mundell-Fleming is that states could select only two items from the following menu: free capital flows, fixed exchange rates, and autonomous monetary policy.[24] Thus it was assumed that by abandoning fixed rates, states could pursue the monetary policies of their choice. But this did not turn out to be the case. In theory, given flexible rates, policy disparities should be mediated at the border. If a state's policies resulted in an inflation rate that was 5 percent above the international average, its currency would depreciate by 5 percent and that would be that. But if capital mobility is accompanied by a consensus with regard to what is a "correct" monetary policy, then the depreciation will not stop at 5 percent. Capital flight in this case will punish the state for pursuing a deviant policy. Failure to reverse that policy in the wake of a sustained depreciation will stimulate even further capital flight and depreciation, and so on. Thus a state that preferred to pursue a more expansionist monetary policy than average, even one that was willing to tolerate the depreciation necessary to restore equilibrium in international prices, may be unable to chart such a course in the face of punishing (as opposed to equilibrating) capital flows. Thus the "holy trinity" is a myth. If there is an ideological consensus regarding macroeconomic policy, it is not possible to have capital mobility and policy autonomy at the same time.

Why does this matter? Because macroeconomic policies are often cloaked in the veil of economic legitimacy. Legitimate policies thrive, it is argued, because they are the most efficient. Those that deviate lose out in a Darwinian struggle—they fail because they are unable to deliver the goods. But two observations raise fundamental questions about this version of events. First, as Friedman and Schwartz argued with regard to the choice of a monetary standard, some policies, perhaps even the best policies, may be unsustainable solely because people (erroneously) think they are inefficient. Thus "legitimacy" does not inevitably yield optimality but is instead a path-dependent self-fulfilling prophecy.[25] Second, close examination reveals that for many macro-

[24] Peter B. Kenen, "Macroeconomic Theory and Policy: How the Closed Economy Was Opened," in Ronald W. Jones and Peter B. Kenen, eds., *Handbook of International Economics*, vol. 2 (Amsterdam: North-Holland, 1985); J. M. Fleming, "Domestic Financial Policies under Fixed and under Floating Exchange Rates," *IMF Staff Papers* 9 (November 1962); Robert Mundell, "The Monetary Dynamics of International Adjustment under Fixed Capital," *Quarterly Journal of Economics* 74 (May 1960).

[25] This does not mean, it should be added immediately and emphatically, that any policy is just as good as another. Rather, the argument is solely that there is a plausible set of macroeconomic policies that would provide roughly similar aggregate economic performance and that should be sustainable based on their internal logic.

economic policy choices, the aggregate economic benefits of choosing one course of action over another are ambiguous, modest, and dwarfed by their political and differential effects.[26]

Ideas matter, then, because they can be crucial in determining which policies from the plausible set will be coronated. The aggregate economic consequences of this choice are not likely to be large. But political consequences—which groups, sectors, regions, and so on, are better off with one policy as opposed to another—may be profound. Closer attention to the relationship between ideas and interests is of course in order. Even more importantly, this suggests that students of the *politics* of money need to reconsider their conceptualization of macroeconomics. In particular, students of politics need to place greater emphasis on the differential as opposed to aggregate effects of macroeconomic phenomena. This will require renewed attention to avenues of inquiry in macroeconomic theory that are not emphasized in contemporary economics.

MACROECONOMICS, MICROPOLITICS

Modern macroeconomic theory assumes aggregate monetary neutrality, that monetary phenomena do not affect the overall performance of the real economy in the long run. This is not an unreasonable assumption for a discipline concerned with long-run aggregate economic growth.[27] But it provides an inappropriate point of departure for most inquiries into politics. Because even if it is true that monetary phenomena do not affect aggregate economic performance in the long run, they do have real distributional economic effects: shaping relative prices, income distribution, and the relative performance of different sectors and industries across the economy.

[26]As Cohen notes, economic efficiency does not explain the most basic contours of the geography of money. While there are theories of "optimal currency areas," it is clear that the use of money is defined by political concerns. Externally the politics of monetary integration cannot be divorced from international politics. Internally the aggregate benefits of such measures, such as the reduction in transaction costs are surely dwarfed by the differential effect that they will have on various groups. See Cohen, *Geography*, 83–84; also Barry Eichengreen and Jeffry Frieden, "The Political Economy of European Monetary Unification," *Economics and Politics* 5, no. 2 (1993). Similarly, it should also be noted that the choice of the adjustment mechanism is made from a range of options with minor differences in their relative efficiency but profound variation in their political consequence. See, for example, Beth Simmons, *Who Adjusts?* (Princeton: Princeton University Press, 1994).

[27]On monetary neutrality, see Robert Lucas, "Nobel Lecture: Monetary Neutrality," *Journal of Political Economy* 104 (August 1996); Milton Friedman, "The Quantity Theory of Money," and Don Patinkin, "Neutrality of Money," both in John Eatwell, Murry Milgate, and Peter Newman, eds., *The New Palgrave: Money* (New York: Macmillan, 1989); Frank Hahn, *Money and Inflation* (Cambridge: MIT Press, 1983), esp. 38–42, 61; Douglas Gale, *Money: In Equilibrium* (Cambridge: Cambridge University Press, 1982), esp. 13–15, 53.

The effect of monetary phenomena on relative prices has been recognized even by those economists who were passionate defenders of aggregate monetary neutrality. Irving Fisher, the patriarch of American monetarism, described his opus, *The Purchasing Power of Money*, as "at bottom simply a restatement of the old 'quantity theory of money.'" Yet he also argued that "prices do not, and in fact cannot, move in perfect unison." Rather, aggregate monetary neutrality holds because there exists "a compensation in price movements in the sense that the failure of one set of prices to respond to any influence on the price level will necessitate a correspondingly greater change in other prices."[28] Fisher (and other macroeconomists), then, did not deny the existence of distributional effects; rather, those effects were simply not the object of their study.

Contemporary macroeconomic theory imposes strict assumptions of homogeneity on its models, and these eliminate the issue of distribution from the study of monetary phenomena.[29] This was not always the case. Classical economists, including "virtually all" of the quantity theorists of the late nineteenth century, believed that monetary phenomena have significant real consequences.[30] The main student of these phenomena was Richard Cantillon, whose eighteenth-century writings produced a research agenda that was sustained until the 1930s. Cantillon argued that changes in the money supply did not descend uniformly from heaven but had to flow from some particular source. The nature of the source and the recipients of the money would determine the distinct effect of a change in the money supply. Following a monetary expansion, the price level will rise, but not uniformly, due to the different tastes and elasticities of those who receive the new money first.[31]

Whereas Cantillon says that "everybody agrees that the abundance of money or its increase in exchange raises the price of everything." He adds, crucially, that "market prices will rise more for certain things than

[28] Irving Fisher, *The Purchasing Power of Money*, rev. ed. (New York: Macmillan, 1920), vii, 196, 197.

[29] See, for example, Olivier Blanchard and Stanley Fischer, *Lectures on Macroeconomics* (Cambridge: MIT Press, 1989).

[30] David Laider, *The Golden Age of the Quantity Theory of Money* (Princeton: Princeton University Press, 1991), 17. Malthus, for example, argued that it "must always be recollected" that real effects of monetary expansion are the result not of "the *quantity* of the circulating medium . . . but the *different distribution* of it." T. R. Malthus, "Depreciation of Paper Money," *Edinburgh Review* 17, no. 34 (February 1811), in Bernard Semmel, ed., *Occasional Papers of T. R. Malthus* (New York: Burt Franklin, 1963), 96.

[31] Michael Bordo, "Some Aspects of the Monetary Economics of Richard Cantillon," *Journal of Monetary Economics* 12 (August 1983), esp. 242–43; Douglas Vickers, *Studies in the Theory of Money, 1690–1776* (Philadelphia: Chilton Company, 1959), esp. 189, 206, 208. For more on Cantillon, see Antoin E. Murphy, *Richard Cantillon: Entrepreneur and Economist* (Oxford: Clarendon Press, 1986); Henry Higgs, "Richard Cantillon," *Economic Journal* 1 (June 1891); and also Joseph Spengler, "Richard Cantillon: First of the Moderns," *Journal of Political Economy* 62 (August 1954) and 62 (October 1954).

for others however abundant the money may be." There are real distributional effects as monetary phenomena work their way through the economy, just as "a river which runs and winds about in its bed will not flow with double the speed when the amount of water is doubled."[32]

Cantillon's work was followed in the nineteenth century by John E. Cairnes and in the twentieth by John Maynard Keynes.[33] Keynes, in his *Treatise on Money*, followed Cantillon and Cairnes with his argument that changes in purchasing power that result from monetary contractions or expansions are "not spread evenly or proportionately over the various buyers." The "new distribution of purchasing power" will have "social and economic consequences" that "may have a fairly large lasting effect on relative price levels." Once again, the argument here is not that monetary disturbances affect aggregate output; rather, it emphasizes the consequences of monetary phenomena even in the absence of such effects. "The fact that monetary changes do not affect all prices in the same way, to the same degree, or at the same time, is what makes them significant."[34]

Ironically, it was subsequent work by Keynes that led to the atrophy of this line of inquiry. In *The General Theory* (1936), which dominated the profession for a generation and ushered in the era of modern macroeconomics, Keynes shifted his emphasis away from monetary influences. The vigorous monetarist response that followed kept the arena of the debate where it was, at the aggregate level.[35] But students of the politics of money need to revisit and build upon the insights of this literature, which was not superseded but rather was left to atrophy, as the focus of the economics profession followed an alternative trajectory. It is the differential, political effects of macroeconomic phenomena that

[32] Richard Cantillon, *Essai sur la nature du commerce en général*, ed. and trans. Henry Higgs (1931; New York: Augustus Kelley, 1964), 161, 179, 177. For practical illustrations of distribution effects, see pp. 163, 165. W. Stanley Jevons characterized Cantillon's tracing of monetary disturbances as "marvelous." See Jevons, "Richard Cantillon and the Nationality of Political Economy," *Contemporary Review* 39 (January 1881), 72. It is important to note that Cantillon was no monetary crank; in fact, he was quite orthodox. He did not believe that a monetary expansion would increase economic activity in the long run. Rather, he was an opponent of inflation who believed that the inflationary process would have self-reversing effects with attendant real economic costs. And he was sensitive to the danger that macroeconomic policy might be manipulated by politicians for their own benefit. See Cantillon, p. 323; also Vickers (fn. 31), 212, 216; Bordo (fn. 31), 237, 251; Murphy (fn. 31), 263, 277.

[33] John E. Cairnes, *Essays in Political Economy: Theoretical and Applied* (1873; New York: Augustus Kelley, 1965), esp. 9, 10; also Michael Bordo, "John E. Cairnes on the Effects of the Australian Gold Discoveries, 1851–73," *History of Political Economy* 7 (Fall 1975).

[34] Keynes, *A Treatise on Money I: The Pure Theory of Money* (1930), in Moggridge and Johnson (fn. 14), 5:81, 83–84.

[35] See Perry G. Mehrling, *The Money Interest and the Public Interest: American Monetary Thought, 1920–1970* (Cambridge: Harvard University Press, 1997); and Charles Rist, *History of Monetary and Credit Theory: From John Law to the Present Day*, trans. Jane Degras (1940; New York: Augustus M. Kelley, 1966), esp. 148, 375.

provide the missing link in the relationship between ideas and interests and clarify the extent to which the choice of a "legitimate" economic policy may be of greater political than economic consequence.

This can be illustrated with two policy examples—one regarding inflation management and the other, capital mobility—that are of great practical importance and contemporary salience. In each case, one policy has a sole, consequential, and self-reinforcing claim to economic legitimacy. But in each case there are also alternative policies that from an economic perspective are equally plausible. Despite their internal consistency, however, they are virtually unsustainable solely because they lack legitimacy. What matters here is not the policy choice per se—as just noted, either is plausible—but rather the way in which the winning policy is anointed. Stripped of their rhetorical dressing, the aggregate economic distinctions between competing policies are ambiguous and modest and dwarfed by their differential effects. Narrow political interests better account for the path chosen.

Imagine Low Inflation

A dramatic example of the extent to which economic legitimacy cloaks profound political consequences can be seen with the hegemonic proposition that vigilance against the threat of inflation must be the primary, if not the sole, goal of macroeconomic policy. This view derives from the recognition, supported by evidence, that governments cannot call forth greater employment and production in the long run via monetary expansion. But the only conclusion that can be drawn from this is that governments should not purposefully promote inflation to expand output. It does not address the question of whether inflation itself is costly or "bad" or whether suppressing inflationary embers need be the principal goal of policy.

There are many deductive reasons why inflation might impose real economic costs, and models can be constructed to simulate why inflation should be costly in practice. Inflation, for example, can weaken the informational role of prices. This can also reduce efficiency by increasing uncertainty. There is empirical support for these propositions, as inflation has been shown to be associated with greater variability in prices and the variability of real output. Inflation, as a tax on cash balances, also carries with it the inefficiencies associated with excise taxes.[36]

[36] For a summary of possible costs, see Stanley Fischer and Franco Modigliani, "Toward Understanding of the Real Effects and Costs of Inflation," in Fischer, *Indexing, Inflation, and Economic*

There are, however, competing arguments that provide reasons why some inflation might actually be a good thing (that is, the inflation rate associated with the maximum possible rate of economic growth is positive). Such arguments often center around the view that in an economy where nominal prices are sluggish to adjust downward, some inflation would allow changes in relative prices, an essential feature of any growing economy, to occur faster and more efficiently. In fact, "inflation may be a necessary part of the process," and deductive models can be constructed and simulations run to show that in fact moderate inflation "permits maximum employment and output."[37]

Given the indeterminacy of deductive arguments and the relative abundance of data on inflation rates and growth rates, the obvious next step is to evaluate the evidence. In fact, the costs of moderate inflation are extraordinarily difficult to find. Inflation hawk Robert Barro, who had previously written that while economists assume that inflation is costly, they "have not presented very convincing arguments to explain these costs,"[38] attempted to illustrate those costs in two recent papers. He concludes that the data reveal costs that seem small yet have a cumulative effect that is "more than enough to justify a strong interest in price stability." He notes, however, that the "clear evidence" of the costs of inflation come from countries that have had inflationary episodes exceeding 10–20 percent per year. The qualifications are even stronger in his follow-up paper: "For inflation rates below twenty percent per year . . . the relation between growth and inflation is not statistically significant."[39]

Policy (Cambridge: MIT Press, 1986). On information, see Gardner Ackley, "The Costs of Inflation," *American Economic Review* 68 (May 1978). On variability, see Richard W. Parks, "Inflation and Relative Price Variability," *Journal of Political Economy* 86 (February 1978). On cash-balance effects, see Martin J. Bailey, "The Welfare Cost of Inflationary Finance," *Journal of Political Economy* 64 (April 1956). Deductive models that can yield high costs of inflation include Michel Dotsey and Peter Richmond, "The Welfare Cost of Inflation in General Equilibrium," *Journal of Monetary Economics* 37 (February 1996); and Martin Feldstein, "The Welfare Cost of Permanent Inflation and Optimal Short-Run Economic Policy," *Journal of Political Economy* 87, no. 4 (1979).

[37] James Duesenberry, "Inflation and Income Distribution," in Eric Lundberg, ed., *Inflation Theory and Anti-Inflation Policy* (Boulder, Colo.: Westview Press, 1977), 265 (first quote); George Akerlof, William Dickens, and George Perry, *The Macroeconomics of Low Inflation*, Brookings Papers on Economic Activity, no. 1 (1996), 2 (second quote). Very low inflation also might undermine monetary policy, given a nominal interest-rate floor of 0 percent.

[38] Robert J. Barro and David B. Gordon, "Rules, Discretion, and Reputation in a Model of Monetary Policy," *Journal of Monetary Economics* 12 (February 1983), 104.

[39] Robert J. Barro, "Inflation and Economic Growth," *Bank of England Quarterly Bulletin* 35 (May 1995), 1, 9; and idem, "Inflation and Growth," *Federal Reserve Bank of St. Louis Review* 78 (May–June 1996), 159. Critics of these papers have challenged Barro's policy prescriptions. W. Stanners argues that even the "weak conclusion" of the first paper "cannot be sustained." Stanners, "Inflation and Growth," *Cambridge Journal of Economics* 20 (July 1996), 511, also 512. In commentary following the second paper, Kocherlakota Narayana argues that "I would recommend that policymakers *not* view lower long run growth as a penalty of inflationary monetary policy" (p. 172).

Almost all of the negative relationships found between inflation and growth are dependent on the consequences of very high levels of inflation.[40] Ultimately, any real economic costs of inflation, especially inflation below 20 percent and certainly below 10 percent, are almost impossible to show. While one study has claimed to demonstrate that inflation over 8 percent is costly (with inflation below 8 percent having a slight positive effect), another study claims to produce "direct evidence against the view that inflation and output growth are reliably related in the long run."[41] But even studies that manage to show some relationship between inflation and growth need to be interpreted very cautiously.[42]

For those who study the politics of money, the final resolution of this debate is not of great concern. The evidence overwhelmingly supports the view that inflation rates at low or moderate levels have very little effect on the performance of the aggregate economy. This casts new light on the long-held understanding that inflation has distributional consequences[43] and on the less appreciated but just as important observation

[40] Two prominent examples of this would be Stanley Fischer, "The Role of Macroeconomic Factors in Growth," *Journal of Monetary Economics* 32 (December 1993); and Michael Bruno, "Does Inflation Really Lower Growth?" *Finance and Development* 32 (September 1995), 35, 38. Yet Bruno and Fischer each support policies designed to keep inflation very low. See Fischer, "Maintaining Price Stability," *Finance and Development* 33 (December 1996), 34–37; Michael Bruno and William Easterly, "Inflation and Growth: In Search of a Stable Relationship," *Federal Reserve Bank of St. Louis Review* 78 (May–June 1996), 145.

[41] Michael Sarel, "Non-linear Effects of Inflation on Economic Growth," *IMF Staff Papers* 43 (March 1996); James Bullard and John W. Keating, "The Long-Run Relationship between Inflation and Output in Postwar Economies," *Journal of Monetary Economics* 36 (December 1995), 495. Bullard and Keating also note some positive effects: "To the extent that we do find statistically significant estimates, they tend to be positive, with a permanent increase in inflation being associated with a permanent increase in the level of output" (p. 494).

[42] First, the findings of studies showing relationships between inflation (and other macroeconomic variables) and growth are quite fragile. See Ross Levine and David Renelt, "A Sensitivity Analysis of Cross-Country Growth Regressions," *American Economic Review* 82 (September 1992). Second, inflation, especially high inflation, might be associated with lower economic performance because high inflation might be a symptom of government incompetence. In this case, the inflation could be a symptom of government-inhibited growth rather than a cause. This is recognized by Fischer (fn. 40, 1993), 487. Third, as Bruno notes (fn. 40), an association of inflation with growth could be the result of other factors, such as supply shocks, that would affect both factors simultaneously (p. 35).

[43] Few would dispute the notion, for example, that unanticipated inflation benefits debtors at the expense of creditors or that governments often use inflation as a means of increasing their resources relative to society. In fact, inflation affects distribution through a multitude of channels, with differential effects on various individuals, classes, sectors, and regions.

Empirical investigations into the effects of inflation include Edward Wolff, "The Distributional Effects of the 1969–75 Inflation on Holdings of Household Wealth in the United States," *Review of Income and Wealth* 25 (June 1979); G. L. Bach and James Stephenson, "Inflation and the Redistribution of Wealth," *Review of Economics and Statistics* 61 (February 1974); William D. Nordhaus, "The Effects of Inflation on the Distribution of Economic Welfare," *Journal of Money, Credit, and Banking* 5 (February 1973); Edward C. Budd and David F. Seiders, "The Impact of Inflation on the Distribution of Income and Wealth," *American Economic Review* 61 (May 1971); Armen A. Alchian and Reuben A. Kessel, "Redistribution of Wealth through Inflation," *Science* 130 (September 4, 1959); G. L. Bach and

by Charles Rist, that "price stability, just as much as much as price instability, gives rise to inequalities as between the citizens of one and the same country."[44]

If vigilance against inflation were the unambiguously universal optimal economic policy, then these distributional quibbles would be of only marginal interest—all groups seem to support bad policies that advance their own narrow interests. But as the economics becomes more ambiguous, the demand for a political explanation must increase. If the hypervigilant policy is but one plausible policy available from a larger menu, then it is likely that the distinct distributional effects of each policy choice, not its economic efficiency, explain the outcome.

This perspective can also explain the support for another pillar of legitimacy, the sanctity of central bank independence (CBI). Support for CBI derives from the need to guard against inflation. And increased CBI is, in fact, associated with lower rates of inflation. But while there is evidence that independent central banks are associated with lower inflation, there is no evidence that they are associated with enhanced economic performance.[45] This is surprising, since, for example, it was assumed that the greater credibility of independent central banks would cause disinflationary episodes to be both shorter and less costly. In fact, the opposite is true. "In direct contradiction" to the credibility hypothesis, "disinflation appears to be consistently *more* costly and no more rapid in countries with independent central banks." Rather than receiving a bonus, independent central banks "have to prove their toughness repeatedly, by being tough."[46]

But monetary policy can be too tight, and disinflationary policies are unambiguously associated with output losses. A policy of emphasizing inflation fighting and delegating responsibility for monetary policy to

Albert Ando, "The Redistributional Effects of Inflation," *Review of Economics and Statistics* 39 (February 1957); Reuben A. Kessel, "Inflation Caused Wealth Redistribution: A Test of a Hypothesis," *American Economic Review* 46 (March 1956); Hyman Sardy, "The Economic Impact of Inflation in Urban Areas," and Zbignew Landau, "Inflation in Poland after World War I," both in Neil Schumukler and Edward Marcus, eds., *Inflation through the Ages* (New York: Columbia University Press, 1983).

[44] Rist (fn. 35), 375.

[45] See Alberto Alesina and Lawrence H. Summers, "Central Bank Independence and Macroeconomic Performance: Some Comparative Evidence," *Journal of Money, Credit, and Banking* 25, no. 2 (1993); and Alex Cuckierman, "Central Bank Independence and Monetary Control," *Economic Journal* 104 (1994), 1440. For a good survey of this large literature, see Sylvester Eijffinger and Jakob De Haan, *The Political Economy of Central-Bank Independence*, Special Papers in International Economics 19 (Princeton: International Finance Section, Princeton University, 1996).

[46] Adam Posen, "Central Bank Independence and Disinflationary Credibility: A Missing Link?" *Federal Reserve Bank of New York Staff Reports* 1 (May 1995), 3, 13 (first quotes); Guy Dabelle and Stanley Fischer, "How Independent Should a Central Bank Be," in Jeffrey C. Fuhrer, ed., *Goals, Guidelines and Constraints Facing Monetary Policymakers* (Boston: Federal Reserve Bank of Boston, 1995), 205 (last quote); see also Eijffinger and De Haan (fn. 45), esp. 36, 38, 64.

independent central bankers devoted to crushing inflation risks disregarding this potential threat to the real side of the economy.[47] In any event, the aggregate economic effects of the sole "legitimate" policy choice (CBI) are modest and ambiguous.

A focus on the real, microlevel effects of monetary phenomena clarifies the sharp politics, as opposed to the ambiguous economics, that is driving macroeconomic policy. Low inflation and perhaps more importantly the macroeconomic policies designed to assure that it remains very low benefit some groups in society at the expense of others. As Adam Posen has argued, low inflation is only sustainable if it has adequate political support. That support, he suggests, comes from numerous sources but especially from "one historically prominent interest group: the financial sector." The financial sector supports central bank independence "as a long-run means to price stability" and it could not be sustained "without that group's ongoing protection of its counter-inflationary activities."[48]

The idea that any threat of inflation must be suppressed is the only one that is legitimate in contemporary mainstream economic theory—and practice. That consensus helps sustain low inflation policies and, more importantly, serves to undermine policies that might deviate from the norm. The evidence supports the contention, however, that the aggregate economic consequences of most levels of inflation are modest, ambiguous, and certainly dwarfed by its differential effects.[49]

Capital Mobility: In Whose Interest?

If inflation fighting dominates macroeconomic policy at the domestic level, the deregulation of capital flows tops the international agenda. Once again, however, the same combination can be observed—self-

[47] It bears repeating that the ultimate resolution of the economic debate over optimal inflation policy is of small concern to students of politics, given the modest stakes. The purpose of this discussion is not to champion one policy over another but rather to illustrate that more than one policy is theoretically plausible, that the perception of legitimacy can be a crucial factor in determining which policy is chosen, and that the differential effects of each easily outweigh the difference in their aggregate economic consequences.

[48] Posen also shows a correlation between the strength of the financial sector and the degree of central bank independence. See Adam Posen, "Declarations Are Not Enough: Financial Sector Sources of Central Bank Independence," *NBER Macroeconomics Annual* 10 (1995), 254, 256, 264; idem, "Why Central Bank Independence Does Not Cause Low Inflation: There Is No Institutional Fix For Politics," in Richard O'Brien, ed., *Finance and the International Economy* 7 (1993), esp. 48.

[49] Studies that have explored the differential effects of low inflation and tight monetary policies include G. J. Santoni, "The Effects of Inflation on Commercial Banks," *Federal Reserve Bank of St. Louis Review* (March 1986); T. F. Cargill and M. M. Hutchison, "The Federal Reserve and the Bank of Japan," in Thomas Mayer, ed., *The Political Economy of American Monetary Policy* (Cambridge: Cambridge University Press, 1990), 172–73; Gerald Epstein and Juliet Schor, "Corporate Profitability as a Determinant of Restrictive Monetary Policy: Estimates for the Postwar United States," in Mayer;

fulfilling legitimacy, ambiguous economics, and underappreciated politics. And if other economic policies—in this instance, some regulation of capital flows—are sound, plausible, and sustainable from an economic perspective (absent problems arising solely from the perception of legitimacy), then an explanation for the choice of one policy over another must be rooted in political analysis, rather than in economic theory.

The idea that capital flows should not be regulated is more than just self-fulfilling (though that is of course profoundly consequential); it is also the express policy of both powerful states like the United States and the international institutions that are supposed to oversee the smooth functioning of the global economy. The support for and the salience of the idea of freeing capital gathered momentum in the 1990s. In the wake of this trend, the International Monetary Fund embarked upon a fundamental revision of its charter, announcing plans in May 1997 to amend its constitution—the Articles of Agreement—"to make the promotion of capital account liberalization a specific purpose of the IMF and give it jurisdiction over capital movements."[50] This would be a dramatic change—in fact, the very *opposite* of what the founding fathers of the IMF intended. They thought that capital controls were necessary to assure the smooth functioning of an open international economy. Thus, the Bretton Woods era, the "golden age of capitalism," was a period of ubiquitous capital control. Now, however, the IMF has asserted that capital liberalization is the only legitimate path to economic efficiency, and it has explicitly proclaimed that "forces of globalization must be embraced." Its new policy has been repeatedly characterized as a proposition "to make unrestricted capital flows a condition of membership in the global economy."[51]

As with very low inflation, there are good deductive reasons to believe in the elimination of capital controls. Openness to capital inflows expands the resources available to the local economy, and the elimination of restrictions on capital outflows gives foreigners the confidence to invest. Not only are investors confident that they will be able to repatriate their profits, but states that allow unrestricted capital flows enjoy greater credibility: market actors assume that they are more likely to

Jeffry Frieden, "Monetary Populism in Nineteenth Century America: An Open Economy Interpretation," *Journal of Economic History* 57 (June 1997).

[50] "IMF Wins Mandate to Cover Capital Accounts," *IMF Survey* (May 12, 1997), 131–32.

[51] "Forces of Globalization Must Be Embraced," *IMF Survey* (May 26, 1997), 131; Darren Mcdermott and Leslie Lopez, "Malaysia Imposes Sweeping Currency Controls: Such Capital Restrictions Win Credence in Wake of Financial Turmoil," *Wall Street Journal*, September 2, 1998; G. Pierre Goad, "Acceptance of Capital Controls Is Spreading," *Asian Wall Street Journal*, September 2, 1998, "condition of membership" quotes.

follow "sound" policies, because if they failed to do so, they would be subject to hemorrhaging of both foreign and domestic capital. More generally, free capital seems to follow the logic of free trade—few deny that an open market leads to goods results in a more efficient allocation of resources, expanded consumption choices, and a host of other benefits such as the discipline imposed by international competition.

But again, as with inflation, there are competing deductive arguments that suggest that some positive level of capital control is optimal for achieving economic efficiency.[52] The free flow of capital differs in important ways from the free flow of goods, just like Cohen's monetary oligopolists differ from oligopoly producers in the real economy. Two attributes make capital quite distinct from most real goods. First, contemporary technology allows investors to move huge amounts of money almost instantaneously and at very little cost. Second, to an important extent, financial assets are worth what people think they are worth. Given these elements, fears regarding what other people are thinking can cause herding behavior, unleashing financial stampedes with economic consequences that veer far from the path suggested by any reading of the economic "fundamentals."

Additionally, in a world of perfectly mobile capital, investors can scan the globe for the best rates of return, and this creates pressure for conformity across countries' macroeconomic policies. But it is highly unlikely that all states should be pursuing the same macroeconomic policies at any given moment. On the contrary, because states face diverse economic conditions, they need to tailor their economic policies accordingly. But without any restrictions on capital, governments that deviate from the international norm, even when pursuing policies appropriate for local needs, are "punished" by capital flight and are often forced to abandon or even reverse such policies.

In this instance, competing deductive arguments have not been followed by a trove of empirical studies ready for mining. Jagdish Bhagwati, noted champion of free trade, recently took many of his fellow economists to task for simply assuming the case for unregulated capital. Proponents of free trade, he observes, have provided mountains of evidence to support their claims; the supporters of free capital have not. If fact, he concludes, "the weight of evidence and the force of logic point

[52] Once again, it is important to note the qualified nature of this argument. The competing argument is not that capital flows are bad but rather that completely deregulated capital would lead to a suboptimally high level of flows.

in the opposite direction, toward restraints on capital flows."[53] This challenge is only reinforced by the recent study of Dani Rodrik, whose analysis from a sample of one hundred countries finds "no evidence that countries without capital controls have grown faster, invested more, or experienced lower inflation."[54]

Skepticism about the benefits of unlimited capital mobility would appear to make more sense with the unexpected spread and depth of the Asian financial crisis. Moreover, efforts by many states to defend their currencies in an environment of mobile capital required deflationary measures that only exacerbated economic distress, while countries that had retained their capital controls were largely spared. And the crisis was, to say the least, unanticipated. In September 1996 the IMF asserted that "international capital markets appear to have become more resilient and are less likely to be a source of disturbances."[55] As late as May 1997 the managing director of the IMF remarked that "global economic prospects warranted 'rational exuberance.'" In addition, economic prospects were "bright" and "overheating pressures have abated in many emerging market economies, especially in Asia—where growth has stayed strong for several years."[56]

But the IMF's retrospective analyses of the crisis make clear that it has not been shaken in its beliefs but remains focused on the domestic sources of the crisis and highly suspicious of any forms of capital control.[57] However sincerely those ideas may be held, given the muddled message yielded by the economic arguments, greater attention must be

[53] Jagdish Bhagwati, "The Capital Myth," *Foreign Affairs* 77 (May–June 1998), 9, 12. For further skepticism and qualifications, see Richard N. Cooper, "Should Capital Controls Be Banished?" *Brookings Papers on Economic Activity* 99:1 (1999).

[54] Dani Rodrik, "Who Needs Capital Account Convertibility?" in *Should the IMF Pursue Capital Account Convertibility?* Essays in International Finance, no. 207 (Princeton: International Finance Section, Princeton University, May 1998), 61.

[55] *IMF Survey* (September 23, 1996), 294. Under the headline "International Capital Markets Charting a Steadier Course," the fund also noted that "although the scale of financial activity continues to grow, market participants—including high-risk high-return investment funds—are more disciplined, cautious, and sensitive to market fundamentals" (p. 293).

[56] *IMF Survey* (May 12, 1997), 129–30.

[57] International Monetary Fund, *International Capital Markets: Developments, Prospects, and Key Policy Issues* (Washington, D.C.: IMF, September 1998), esp. 6, 11, 57, 63, 73, 148–50; see also International Monetary Fund, *World Economic Outlook. Financial Turbulence and the World Economy* (Washington, D.C.: IMF, October 1998), esp. 6–18, 101–2. It should be noted, however, that in the wake of the crisis the World Bank has been willing at least to address the issue of the possible benefits of some control over short-term capital flows. See World Bank, *Global Economic Prospects and the Developing Countries, 1998–99: Beyond Financial Crisis* (Washington, D.C.: World Bank, 1999), esp. xi–xii, xxi, 4, 123–24, 128, 142–52; see also World Bank, *East Asia: The Road to Recovery* (Washington, D.C.: World Bank, 1998), esp. 9–10, 16, 34.

paid to the prospect that interest rather than efficiency is behind the drive for financial liberalization. States and state-led institutions are responsible for global financial deregulation.[58] And according to at least one report, there are those in Washington who have "quietly expressed the hope" that Malaysia's response to the financial crisis, the introduction of capital controls, "would fail so spectacularly that the smoldering ruins of the Malaysian economy would act as a caution to other countries."[59] Surely the narrow interests of individual states or influential groups within those states are at stake. The modest and ambiguous aggregate gains to be reaped from perfect global capital mobility cannot account for such passions.

Although they approach the issue from distinct, even competing perspectives, Cohen and Strange agree that the contours of global economy have been transformed by the ascendance of money, and these changes have basic and important political consequences. IPE scholars at the start of the twenty-first century will be called upon increasingly to come to grips with these issues. The resulting studies will undoubtedly trace their points of departure, questions of origin, and analytical frameworks back to either *The Geography of Money* or *Mad Money*, or perhaps even both. They will also necessarily confront the unique interconnections between the ideas, material interests, and institutions associated with the management of money. From these links derive many of the riddles and mysteries that animate these books and that have plagued and energized the study of money throughout history: That money is what people think it is. That apparently technical debates are inescapably political. And that those politics, more than the underlying economics, can best explain the shape of the monetary landscape.

[58] See Eric Helleiner, *States and the Reemergence of Global Finance* (Ithaca, N.Y.: Cornell University Press, 1994).

[59] David E. Sanger, "Gaining Currency: The Invisible Hand's New Strong Arm," *New York Times*, September 9, 1998.

ABSTRACTS

Partition as a Solution to Ethnic War

AN EMPIRICAL CRITIQUE OF THE THEORETICAL LITERATURE

By NICHOLAS SAMBANIS

Theorists of ethnic conflict have argued that the physical separation of warring ethnic groups may be the only possible solution to civil war. They argue that without territorial partition and, if necessary, forced population movements the war cannot end and genocide is likely. Other scholars have counterargued that partition only replaces internal war with international war, that it creates undemocratic successor states, and that it generates tremendous human suffering. This debate has so far been informed by very few important case studies. This article uses a new data set on civil wars to identify the main determinants of war-related partitions and estimate their impact on democratization, on the probability that war will recur, and on low-level ethnic violence. This is the first large-N quantitative analysis of this topic, testing the propositions of partition theory and weighing heavily on the side of its critics. Most assertions of partition theorists fail to pass rigorous empirical tests. The article also identifies some determinants of democratization after civil war, as well as the determinants of recurring ethnic violence. These empirical findings are used to formulate an alternative proposal for ending ethnic violence.

Degrees of Democracy

SOME COMPARATIVE LESSONS FROM INDIA

By PATRICK HELLER

This article draws on the case of India to address the question of democratization by exploring the dynamic interplay of the formal, effective, and substantive dimensions of democracy. Fifty-three years of almost uninterrupted democratic rule in India have done little to reduce the political, social, and economic marginalization of India's popular classes. Within India the state of Kerala stands out as an exception. Democratic institutions have effectively managed social conflict and have also helped secure substantive gains for subordinate classes. Kerala's departure from the national trajectory is located in historical patterns of social mobilization that coalesced around lower-class interests and produced forms of state-society engagement conducive to democratic deepening. Contrary to much of the transition literature, this case suggests that high levels of mobilization and redistributive demands have democracy-enhancing effects.

The Ideological Determinants of Liberal Economic Reform

THE CASE OF PRIVATIZATION

By HILARY APPEL

The empirical literature on mass privatization in the postcommunist context emphasizes the preferences and power of interest groups in order to account for the design of privatization. This approach has been consistent with mainstream theories of property rights formation that focus on the self-interested, rationally calculated pursuit of wealth and/or power as the motivation behind the development of new ownership arrangements. Absent from these theories, however, are the ideological and cognitive components in the creation of property rights systems. This lacuna is extremely problematic when considering the postcommunist privatization experience in which specific ideologies—such as anticommunism, liberalism, pro- or anti-Westernism, and nationalism—have profoundly influenced the particular form that new property institutions have taken. This article explores how ideology interacts with the distribution of power and the formation of material interests in society. After considering the shortcomings of strictly material-based theories of property regime change, the article suggests four mechanisms by which ideology determines the design and implementation of privatization programs in postcommunist countries.

What's So Different about a Counterfactual?

By RICHARD NED LEBOW

The author contends that the difference between so-called factual and counterfactual arguments is greatly exaggerated; it is one of degree, not of kind. Both arguments ultimately rest on the quality of their assumptions, the chain of logic linking causes to outcomes, and their consistency with available evidence. He critiques two recent historical works that make extensive use of counterfactuals and finds them seriously deficient in method and argument. He then reviews the criteria for counterfactual experimentation proposed by social scientists who have addressed this problem and finds many of their criteria unrealistic and overly restrictive. The methods of counterfactual experimentation need to be commensurate with the purposes for which it is used. The author discusses three uses for counterfactual arguments and thought experiments and proposes eight criteria appropriate to plausible-world counterfactuals.

Understanding China's Reform

LOOKING BEYOND NEOCLASSICAL EXPLANATIONS

By SHU-YUN MA

The relative success of post-Mao reform in China has raised a number of questions regarding the neoclassical perspective: How could China have achieved rapid economic growth without privatization? Why have Chinese officials not been resistant to market reform? What makes the Chinese state developmental rather than predatory? The four recent works reviewed in this article attempt to answer these questions by moving away from the neoclassical approach, yet none offers a better alternative. The search for the secret of China's economic "miracle" must be continued.

REFEREES

The editorial committee of *World Politics* thanks the following scholars for their assistance in refereeing submissions in 1999:

Steinar Andresen
David Bachman
Michael Barnett
Larry Bartels
Gary Bass
Robert Bates
Charles Beitz
Thomas Berger
Adam Berinsky
Mark Blyth
Michael Bratton
Nathan J. Brown
Valerie Bunce
Jeffrey T. Checkel
Dan Chirot
William R. Clark
Benjamin J. Cohen
Timothy Colton
Giuseppe Di Palma
Avinash Dixit
Richard F. Doner
George W. Downs
Barry Eichengreen
Laura Enriquez
John Entelis
M. Steven Fish
Volker K. Frank
Matt Gabel
Geoffrey M. Garrett
Barbara Geddes
Charles L. Glaser
Ellis Goldberg
Judith Goldstein
James Gow
Fred Greenstein
Merilee Grindle
Lowell W. Gudmundson
Mauro F. Guillen
Stephan Haggard
Stephen E. Hanson
Eric Hershberg
Evelyne Huber
Paul Huth
Christian Joppke
Junko Kato
Mary F. Katzenstein
Peter Katzenstein
Michael D. Kennedy
Marc Kilgour
Herbert Kitschelt
Robert Klitgaard
Audie Klotz
Jeffrey Kopstein
John Langlois
Jeffrey Legro
Jeffrey Lewis
John Lewis
Claudio Lomnitz
John B. Londregan
Ian S. Lustick
Ellen Lust-Okar
James Mahoney
Edward Mansfield
Anthony W. Marx
Walter Mattli
Margaret A. McKean
Joel S. Migdal
Andrew Moravcsik
James D. Morrow
Robert G. Moser
Gerardo Munck
Andrew Nathan
Richard Newfarmer
Kevin J. O'Brien

William E. Odom
Christina Paxson
Minxin Pei
T. J. Pempel
Adam Przeworski
William B. Quandt
Karen Remmer
David Rocke
Philip G. Roeder
Thomas Romer
Gideon Rose
Howard Rosenthal
Donald Rothchild
Todd Sandler
Anne Sartori
Eduardo Silva
Regina Smyth
Jack Snyder
Richard Snyder
Steven L. Solnick
Stephen Stedman
Karen Stenner
M. Scott Taylor
Kathleen Thelen
Daniel S. Treisman
Steven Vogel
Robert Wade
Barbara Walter
John Waterbury
David L. Weimer
Barry Weingast
Bruce Western
Mark Williams
William C. Wohlforth
Oran R. Young
I. William Zartman
Kimberly Zisk

ERRATA

'he following table replaces Table 1 in Judith S. Kullberg and William ;immerman, "Liberal Elites, Socialist Masses, and Problems of ussian Democracy," *World Politics* 51 (April 1999), 336.

TABLE 1
ELITE AND MASS ORIENTATIONS TO THE POLITICAL ECONOMY
(% AND NUMBER, IN THAT ORDER)

	Distribution Including All Respondents			
	1993 Elite	*1995 Elite*	*Mass Sample, European Russia, 1993*	*Mass Sample Russia,1995*
iberal democrats	74.0	72.8	27.4	21.9
	(148)	(131)	(341)	(621)
larket authoritarians	5.0	5.6	7.2	5.8
	(10)	(10)	(89)	(164)
ocial democrats	6.5	7.8	18.2	16.7
	(13)	(14)	(226)	(473)
ocialist authoritarians	4.5	3.9	14.8	13.7
	(9)	(7)	(184)	(388)
mbivalent	7.0	9.4	16.7	27.4
	(14)	(17)	(207)	(779)
nmobilized	3.0	.6	15.8	14.6
	(6)	(1)	(196)	(414)
	100	100	100	100
Total	(200)	(180)	(1,243)	(2,839)

	Distribution Excluding "Ambivalent" and "Unmobilized" Respondents			
	1993 Elite	*1995 Elite*	*Mass Sample, European Russia, 1993*	*Mass Sample Russia,1995*
iberal democrats	82.2	80.9	40.6	37.7
	(148)	(131)	(341)	(621)
larket authoritarians	5.6	6.2	10.6	10.0
	(10)	(10)	(89)	(164)
ocial democrats	7.2	8.6	26.9	28.7
	(13)	(14)	(226)	(473)
ocialist authoritarians	5.0	4.3	21.9	23.6
	(9)	(7)	(184)	(388)
	100	100	100	100
Total	(180)	(162)	(840)	(1,646)

PARTITION AS A SOLUTION TO ETHNIC WAR

An Empirical Critique of the Theoretical Literature

By NICHOLAS SAMBANIS*

INTRODUCTION: THE THEORETICAL CASE FOR PARTITION

IN two influential articles Chaim Kaufmann elaborated a set of hypotheses on the usefulness of partition as a solution to ethnic civil war, building on the arguments of John Mearsheimer and Stephen Van Evera.[1] Before them, a first wave of theorists had considered the benefits and costs of partition. A prominent theorist, Donald Horowitz, suggested that

> if the short run is so problematical, if the constraints on policy innovation are many, if even grand statements need patchwork readjustment, perhaps it is a mistake to seek accommodation among the antagonists. If it is impossible for groups to live together in a heterogeneous state, perhaps it is better for them to live apart in more than one homogeneous state, even if this necessitates population transfers. Separating the antagonists—partition—is an option increasingly recommended for consideration where groups are territorially concentrated.[2]

It is hard to argue with such a statement. Assuming that "the constraints of policy innovation" and "the short run" can be accurately measured ex ante, it would be easy to recommend partition for some

* I thank Michael Doyle, Jeff Herbst, Chris Paxson, Russell Leng, George Tsebelis, Philippos Savvides, and three anonymous referees for their very useful comments and suggestions. This article is part of a World Bank project on "The Economics of Political and Criminal Violence." The opinions and any errors in this paper are the author's and they do not necessarily reflect the views of the World Bank, its executive directors, or the countries they represent.

[1] Chaim Kaufmann, "Possible and Impossible Solutions to Ethnic Civil Wars," *International Security* 20 (Spring 1996); idem, "When All Else Fails," *International Security* 23 (Fall 1998); John J. Mearsheimer and Stephen Van Evera, "When Peace Means War," *New Republic* (December 1995).

[2] Donald L. Horowitz, *Ethnic Groups in Conflict* (Berkeley: University of California Press, 1985), 588. See also Arend Lijphart, *Democracy in Plural Societies* (New Haven: Yale University Press, 1977), 44–47; Robert A. Dahl, *Polyarchy: Participation and Opposition* (New Haven: Yale University Press, 1971), 121; and Samuel P. Huntington, "Civil Violence and the Process of Development," *Adelphi Paper* no. 83 (London: International Institute for Strategic Studies, 1971), 14. Horowitz also discusses dangers of partition (pp. 588–91).

World Politics 52 (July 2000), 437–83

countries while trying to patch up others. However, neither the first nor the second wave of theorists was able to produce operational criteria for applying the theory consistently across cases.

Despite this lack of operational applicability and clarity, partition theory, with its intuitive appeal, has been shaping scholarly and policy opinion on how to end ethnic civil wars. To help policymakers make informed decisions about the usefulness of partition as a strategy to end civil war, I compiled a new data set of all civil wars in the post–World War II era and used that data set to empirically test the set of hypotheses that constitute partition theory.

I focus on the second wave of partition theorists, who have had the greatest impact on the debate. According to them, ethnic violence implies that civil politics cannot be restored unless "ethnic groups are demographically separated into defensible enclaves. . . . Solutions that aim at restoring multi-ethnic civil politics and at avoiding population transfers—such as power-sharing, state re-building, or identity reconstruction—cannot work because they do nothing to dampen the security dilemma."[3]

The so-called security dilemma lies at the core of partition theory. The dilemma in its purest form arises when one community faces a distrustful other and one's actions to increase one's own security are perceived as threatening the security of others.[4] Posen argues that this dynamic is intensified when the opponents belong to different ethnic groups.[5] Ethnic civil wars, argue partition theorists, are characterized by strong and fixed identities, by weak ideological and strong religious overtones, by the dissemination of tales of atrocities to strengthen mobilization, and by easy recognition of identities and the existence of only limited scope for individual choice. Therefore, once war starts, the theory goes, all members of the group must be mobilized because other ethnic groups will inevitably recognize them as enemies.[6] This inescapable destiny reinforces the dynamics of war and must lead to partition, since "once ethnic groups are mobilized for war, the war cannot end until the populations are separated into defensible, mostly homogeneous regions. . . . Ethnic separation . . . allows . . . cleansing and rescue imperatives [to] disappear; war is no longer mandatory."[7]

[3] Kaufmann (fn. 1, 1996), 137, 139.

[4] See Robert Jervis, "Cooperation under the Security Dilemma," *World Politics* 30 (January 1978). Such suspicion and fear would be supported by actual or perceived state collapse, which transforms the domestic political environment into a near anarchic environment.

[5] Barry Posen, "The Security Dilemma and Ethnic Conflict," *Survival* 35 (Spring 1993).

[6] Kaufmann (fn. 1, 1996), 139–47.

[7] Ibid., 150.

However intuitive that reasoning may be, it is nothing more than a series of unsubstantiated assertions. Beyond a handful of self-selected cases, partition theorists have not presented proof that partition is the only viable and credible solution to ethnic civil war. They have not even proven that partition outperforms other war outcomes in terms of peace-building potential. The theory is indeed plausible under strict assumptions, but are these assumptions realistic?

This paper poses a serious challenge to partition theorists by providing a rigorous test of the theory with a comparison of post–civil war realities in both partitioned and nonpartitioned states. It focuses on countries that have experienced civil war; it does not consider cases of peaceful partition.[8] I begin by summarizing other authors' critiques of partition theory in the next section. I then identify the main determinants of war-related partition and test the three core hypotheses of partition theory: (1) that partitions facilitate postwar democratization; (2) that they prevent war recurrence; and (3) that they significantly reduce residual low-level ethnic violence. My tests lead me to reject the most critical tenets of partition theory. I find that partitions do not help prevent recurrence of ethnic war and that they may not even be necessary to stop low-level ethnic violence. Although it may seem like a clean and easy solution, partition fares no better than other outcomes of ethnic civil war. I turn now to a summary of the debate.

REBUTTALS AND COUNTERREBUTTALS: THE STATE OF THE DEBATE ON PARTITION

The suggestion that populations must be forcibly separated to prevent them from killing each other has inspired both approval and criticism. The most significant criticism is that partition may be too limiting a

[8] Partition theorists also approach the problem in this way: they do not discuss partition as a preventive measure before war occurs but rather analyze it as a strategy to end civil war after it occurs—"when all else fails," as Kaufman (fn. 1, 1998) puts it. Peaceful partitions therefore cannot offer any information on my main research question—war recurrence—since a war is a necessary precondition for war recurrence. My research design is therefore the equivalent of a biostatistician's inquiry into the effects of medical treatment for illness: suffering from that illness is a precondition for inclusion in the study. Studying the relationship between initial war occurrence and partition would be an interesting extension of my study. The research question would have to be reformulated, as would the data set. The dependent variable could no longer be war recurrence or residual violence and one would need a theory of civil war occurrence that included partition as a potentially important determinant of civil war (or civil peace). Such a study would analyze a random sample of countries (or the entire population of countries) and would have to include both countries that experienced war and countries that were at peace. To identify whether partition causes war, one could code a binary variable denoting if the country was partitioned and use it as a regressor in a model of the onset of war. Alternatively, one could estimate two separate regressions on partitioned and nonpartitioned countries and compare the coefficients. In terms of the medical research example above, this study would effectively ask: how does factor x affect one's chances of becoming ill?

solution and that ethnic cooperation may be possible even after civil war, facilitated by both ethnic diffusion and third-party security guarantees.[9] Some say that partition is also too severe a solution, as forced population movements cause tremendous human suffering and violate important human rights.[10] The process of partition may also create undemocratic successor states, which would be likely to repress their residual minorities much as their predecessors did.[11] This is important because successor states will rarely be ethnically homogeneous and may incorporate new ethnic antagonisms.[12] Moreover, partition does not resolve underlying ethnic rivalry, so civil wars that end in partition could be transformed into interstate wars between predecessor and successor states.[13] Finally, endorsing some partitions may encourage partition movements elsewhere, leading to new wars.[14]

The debate between partition theorists and their critics is ongoing, although some of the critiques listed above have been settled or are close to being settled in the literature. Below, I summarize the status of three important arguments.

Successful Ethnic Partitions Do Not Encourage Partition Movements Elsewhere

Critics have argued that support by the international community for partition in a few countries would encourage partitions elsewhere. Kaufman, however, has rebutted this criticism by arguing that the uncertainty and extreme costs of civil war would discourage the initiation of partition movements unless such movements are inevitable for domestic political reasons.[15] That position is partially supported by a set of analyses of the "international spread of ethnic conflict" which suggests that cross-country contagion effects of ethnic partition movements are rare.[16]

[9] On ethnic diffusion cooperation, see Daniel L. Byman, "Divided They Stand: Lessons about Partition from Iraq and Lebanon," *Security Studies* 7 (Autumn 1997). On security guarantees and ethnic war termination, see Barbara F. Walter, "The Critical Barrier to Civil War Settlement," *International Organization* 51 (Summer 1997). Neither Byman nor Walter is a critic of partition theory (Byman in fact supports partition under certain conditions). Some of their arguments, however, can be read as indirect critiques of the theory.

[10] Radha Kumar, "The Troubled History of Partition," *Foreign Affairs* 76 (January–February 1997).

[11] Ibid.; see also Amitai Etzioni, "The Evils of Self-Determination," *Foreign Policy* 89 (Winter 1992–93); and Robert Schaeffer, *Warpaths: The Politics of Partition* (New York: Hill and Wang, 1990).

[12] Byman (fn. 9).

[13] Ibid.; and Schaeffer (fn. 11).

[14] Etzioni (fn. 11); Allen Buchanan "Self-Determination and the Right to Secede," *Journal of International Affairs* 45 (Winter 1992).

[15] Kaufman (fn. 1, 1998).

[16] David A. Lake and Donald Rothchild, eds., *The International Spread of Ethnic Conflicts* (Princeton: Princeton University Press, 1998). More research is needed to fully appreciate the impact of historical examples and of precedential reasoning in ethnic conflict.

Successor States May Also Incorporate Ethnic Conflict

According to partition theorists, the success of partition depends on the demographic reorganization of the new territories and on the absence of militarily significant minorities in the new states. However, successor states in most actual cases of partition are not ethnically pure. Hence, this core premise of partition theory may be unrealistic.[17] To quote Horowitz:

> The linchpin of all the arguments [for partition] is the assumption that the probable outcome of secession and partition will be more homogeneous states and, concomitantly, a lower ethnic conflict level. If the assumption were correct, the conclusion would follow. *But the assumption is wrong: the only thing secession and partition are unlikely to produce is ethnically homogeneous or harmonious states.*[18]

Furthermore, even if successor states were homogeneous, the mobilization perspective of ethnic conflict would suggest that, unless partition is accompanied by regime or leadership reform, there is no guarantee that ethnic groups in successor states will not be mobilized into another war against residual minorities.[19] So again the theory's claims depend critically on unrealistic premises about the ethnic composition and political institutions of successor and predecessor states.

This last point is related to the theory's dependence on the concept of the security dilemma, which ignores the fact that conflict is often due not to the defensive security needs of ethnic groups but rather to the "predatory" goals of their leaders. It is worth noting that even the "father" of the concept of the security dilemma—Robert Jervis—has acknowledged that in most contemporary civil conflicts there are not only security motives but also predatory ones. It follows that partition will not resolve the security dilemma of the partitioned ethnic groups if it exacerbates the "predatory" incentives of predecessor states.[20] Yugoslavia's recurrent wars are a case in point.

[17] See Horowitz (fn. 2), 588–91 and chaps. 2, 6.

[18] Ibid., 589, emphasis added.

[19] See, among others, David A. Lake and Donald Rothchild, "Containing Fear: The Origins and Management of Ethnic Conflict," *International Security* 21 (Fall 1996); V. P. Gagnon, "Ethnic Nationalism and International Conflict: The Case of Serbia," *International Security* 19 (Winter 1995); Rui J. P. de Figueiredo and Barry R. Weingast, "The Rationality of Fear: Political Opportunism and Ethnic Conflict," in Barbara Walter and Jack Snyder, eds., *Civil Wars, Insecurity, and Intervention* (New York: Columbia University Press, 1999).

[20] David Laitin, "Somalia: Civil War and International Intervention," in Walter and Snyder (fn. 19); and Jack Snyder and Robert Jervis "Civil War and the Security Dilemma," in Walter and Snyder (fn. 19), 19–24.

ETHNIC COOPERATION IS POSSIBLE WITHOUT PARTITION

There can be many ways to resolve the security dilemma.[21] Noncooperative game theory identifies a number of conditions under which a mutually beneficial Nash equilibrium can be achieved between parties whose preferences are, first, to cheat their opponent into cooperating while they defect and, second, to mutually cooperate rather than mutually defect—preferences such as those characterizing many ethnic conflicts. Perspectives on international negotiation have also suggested that the parties can cooperate if they negotiate a solution to "delegate to neutral authorities."[22] If negotiation is a viable option, it may be possible to reach an internationally or regionally brokered agreement that addresses the conflict's underlying causes.[23]

The problem with these solutions is that they may not be credible, which reinforces the security dilemma.[24] Thus, argues Walter, civil wars tend not to end in negotiated settlements, and a settlement will hold only when external security guarantees are available.[25] However, it need not follow that all peace agreements and institutional solutions to ethnic conflict are noncredible. Only the warring parties can gauge ex ante whether an institutional framework designed to end the war will be successful because they know their opponents and can estimate the probability that the peace process will fail.[26] Also, partition, which is allegedly credible because it redraws national borders to resolve the minority's security dilemma, is as vulnerable to the credibility argument as any other solution, since only robust external security guarantees can credibly prevent predatory predecessor states from restarting the war against successor states.

What if there is no decisive end to the war (such as a military victory) and ethnic competition persists, threatening the possibility of new violence? The rationalist perspective on war would suggest that the war should have resolved any uncertainty about relative resolve and power that might have led to war in the first place. Thus, miscalculations

[21] I do not develop a theory of ethnic cooperation in this paper. I only summarize relevant theoretical arguments to frame my empirical analysis. Thus, this section is not designed to resolve all doubt about the possibility of ethnic cooperation after civil war.

[22] Snyder and Jervis (fn. 20), 18. On power sharing, see Timothy Sisk, *Power Sharing and International Mediation in Ethnic Conflicts* (Washington, D.C: United States Institute of Peace, 1996).

[23] Lake and Rothschild (fn. 19).

[24] Snyder and Jervis (fn. 20).

[25] In Walter's (fn. 9) argument, the security dilemma depends on an asymmetry of power between the government and rebels. Walter notes that *credible* external security guarantees are effective, though difficult. The difficulty in proving the credibility of the third party's commitment amounts to indirect support for the partition thesis, though only if partition is proven to be more credible and less difficult to implement than a brokered settlement.

[26] De Figuereido and Weingast (fn. 19).

would be less likely after the first war and rational parties would prefer not to start a new war regardless of the first war's outcome.[27] This could change, however, as the parties' relative capabilities change over time. So one way for the international community to enable stable peace is to preserve the military balance that follows the end of the war. An alternative, which could work better in some situations, is to create a regional hegemon responsible for regional peace.[28]

Finally, cooperation among ethnic groups may be possible if ethnic diffusion increases as a result of the war, that is, if the opposite of partition takes place. Byman has suggested that increased ethnic diffusion may mitigate the security dilemma, since it would reduce the probability that a single ethnic group could become politically and militarily dominant. His argument derives from the theoretical literature on international alliances and posits that ethnic "balancing" against threatening groups is both possible and stabilizing.[29] This hypothesis has yet to be tested, but it is relevant to note that a budding political economy literature on civil wars has identified a parabolic relationship between ethnic fragmentation and the probability of civil war; that is, the probability of civil war drops significantly at very high levels of ethnic diversity and it is greatest in ethnically polarized societies, which seems to support Byman's hypothesis.[30] The question that partition theorists raise is slightly different, however: can ethnic diversity reduce the risk of war recurrence after the first war ends? I answer this question in later sections.

TAKING SIDES: NEW DATA AND NEW EMPIRICAL TESTS OF PARTITION THEORY

The four most important questions in partition theory are still unresolved. (1) What are the main determinants of partition? (2) Does par-

[27] The rationalist school is well represented by Geoffrey Blainey, *The Causes of War* (New York: Free Press, 1973); and James Fearon, "Rationalist Explanations for War," *International Organization* 49 (Summer 1995). War should reveal any private information about relative power and resolve, making it less rational for parties to resort to war again rather than to strike a more efficient bargain short of war. That said, we should also consider other explanations of war and weigh them against this argument.

[28] Snyder and Jervis (fn. 20); and Laitin (fn. 20).

[29] Byman (fn. 9). This argument can backfire. Ethnic balancing can also paralyze the state. For such an argument, see Harrison Wagner, "The Causes of Peace," in Roy Licklider, ed., *Stopping the Killing* (New York: New York University Press, 1993). Wagner argues that because military victory results in unitary political systems, it will be more stable than any peace agreement based on ethnic balancing. Indeed, the occurrence of an ethnic war suggests a precedent of failed ethnic balancing. In this paper, I present empirical results about the relationship of ethnicity to postwar violence, but that relationship also demands better theorizing.

[30] See, e.g., Paul Collier, Ibrahim Elbadawi, and Nicholas Sambanis, "How Much War Will We See? Estimating the Probability of Civil War in 161 Countries" (Manuscript, World Bank, February 2000).

tition create democratic or undemocratic states? (3) Does partition prevent war recurrence? And (4) does partition end low-level ethnic violence (that is, violence short of war)?

To answer these questions, I have compiled a new cross-sectional data set of all civil wars since 1944. The unit of observation is a civil war. The analysis focuses on wars that have been over for at least two years at the time of writing, but also included are eight ongoing wars to capture the partition theorists' interest in partition as a way of ending ongoing wars.[31]

A civil war is defined as an armed conflict that has (1) caused more than one thousand deaths; (2) challenged the sovereignty of an internationally recognized state; (3) occurred within the recognized boundaries of that state; (4) involved the state as one of the principal combatants; (5) included rebels with the ability to mount an organized opposition; and (6) involved parties concerned with the prospect of living together in the same political unit after the end of the war.[32] This definition allows me to combine wars from several data sets.[33] Detailed

[31] Dropping those cases did not affect any of the results presented in later sections.

[32] This definition is nearly identical to the definition of a civil war in J. David Singer and Melvin Small, *Correlates of War Project: International and Civil War Data, 1816–1992* (Ann Arbor, Mich.: ICPSR, 1994); idem, *Resort to Arms* (Beverly Hills, Calif.: Sage Publications, 1982); and Roy Licklider, "The Consequences of Negotiated Settlements in Civil Wars, 1945–1993," *American Political Science Review* 89 (September 1995). Unlike them, my coding of wars does not presume one thousand deaths per year, but rather uses the one thousand deaths as the threshold for the entire war. In fact, however, most of my cases have caused one thousand deaths annually. My coding decision was based on the arbitrariness of setting one thousand as the annual death criterion and on the lack of available data on annual deaths in the Correlates of War project. Indeed, the codebook of the ICPSR study, which includes the international and civil war data files for the Correlates of War Project, does not mention an annual death threshold and no annual death data are made available by the authors.

[33] My sources for coding wars include Singer and Small (fn. 32, 1994); Licklider (fn. 32); idem (fn. 29); Peter Wallensteen and Margareta Sollenberg, "Armed Conflicts, Conflict Termination, and Peace Agreements, 1989–1996," *Journal of Peace Research* 34, no. 3 (1997); Daniel C. Esty et al., "The State Failure Project: Early Warning Research for US Foreign Policy Planning," in John L. Davies and Ted Robert Gurr, eds., *Preventive Measures: Building Risk Assessment and Crisis Early Warning Systems* (Boulder, Colo., and Totowa, N.J.: Rowman and Littlefield, 1998); David Mason and Patrick Fett, "How Civil Wars End: A Rational Choice Approach," *Journal of Conflict Resolution* 40 (December 1996); Patrick Regan, "Conditions for Successful Third Party Interventions," *Journal of Conflict Resolution* 40, no. 1 (1996); Walter (fn. 9); SIPRI, *SIPRI Yearbook* (http://editors.sipri.se/pubs/yearb.html); Human Rights Watch, *World Report* (New York and Washington, D.C.: Human Rights Watch, various years). Secondary texts consulted include Robert I. Rotberg, ed., *Burma: Prospects for a Democratic Future* (Washington, D.C.: Brookings Institution, 1998); Martin Stuart-Fox, *A History of Laos* (Cambridge: Cambridge University Press, 1998); David Callahan, *Unwinnable Wars* (New York: Hill and Wang, 1997); John O. Iatrides, "The Doomed Revolution: Communist Insurgency in Postwar Greece," in Licklider (fn. 29); Michael W. Doyle, Robert Orr, and Ian Johnstone, eds., *Keeping the Peace* (Cambridge: Cambridge University Press, 1997); Francis M. Deng, *War of Visions: Conflict of Identities in the Sudan* (Washington, D.C.: Brookings Institution, 1999); David McDowall, *A Modern History of the Kurds* (New York: St. Martin's Press, 1996); and Nicholas Sambanis, "United Nations Peacekeeping in Theory and in Cyprus: New Conceptual Approaches and Interpretations" (Ph.D. diss., Princeton University, 1999). The most important difference between my coding and that of others concerns the periodization of wars. I have broken what is a single observation of war in other data

documentation on my coding and sources for all the variables in the data set is available online.[34]

DEFINITION OF PARTITION

The variable denoting partition, PART, is binary and equals 1 if an event of partition is observed and 0 otherwise. Partition is defined as a war outcome that involves both border adjustment and demographic changes. This is a broad definition that differs slightly from Kaufmann's. To justify a narrower definition of partition, Kaufmann wrote that "we should focus on partition rather than secession . . . to assess whether international intervention reduces or increases the costs of ethnic conflict."[35] He then defined partitions as "separations jointly decided upon by the responsible powers: either agreed between the two sides (and not under pressure of imminent military victory by one side), or imposed on both sides by a stronger third party . . . [and he defined] secessions as new states created by the unilateral action of a rebellious ethnic group."[36]

I do not find the narrow definition convincing or useful, given that the far-reaching implications of partition theory affect secessions and partitions equally in the minds of most policymakers and academics. Moreover, the narrow definition reclassifies as secessions certain cases that Kaufmann originally treated as partitions.[37] Finally, this definition does not justify the inclusion of some of the partitions included on Kaufmann's own list.[38] An example is Cyprus, which Kaufmann correctly—though for the wrong reasons—classifies as a de facto partition.[39] The 1974 partition of Cyprus was neither the outcome of an agreement nor an imposition by a third party, as the narrow definition would have it. Rather, it was the result of military victory by the Turkish side.[40]

Given these problems with the narrow definition, I use the broader definition, combining cases of partition and secession listed in Kauf-

sets into more than one observation; or, conversely, I have collapsed two or more observations in one by uniformly applying this rule: a war is coded as a single observation if the parties and issues are the same, if the war events are not separated by a substantial period of nonviolence, and/or if the parties sign a peace agreement or agree to a major truce.

[34] The document can be downloaded from http://www.worldbank.org/research/conflict/data.htm.

[35] Kaufmann (fn. 1, 1998), 125.

[36] See Kaufmann (fn. 1, 1998), 125, fn. 21.

[37] Kaufmann (fn. 1, 1996).

[38] Kauffman (fn. 1, 1998).

[39] Ibid.

[40] It is well known to scholars of the Cyprus problem that Turkey and the Turkish Cypriots constituted and acted as a single party both during the violent part of that conflict (1963–74) and during the subsequent negotiation phases; see Sambanis (fn. 33).

mann's two articles.[41] I also add cases that satisfy my definition but are not on Kaufmann's list.[42] Table 1 lists all civil wars and partitions in my data set, sorted by country name, war start/end dates, the type of war, war recurrence, and lower-level violence outcomes.[43]

In my analysis of democratization, war termination, and low-level political violence I use explanatory variables that other authors have identified as significant for those events. These variables are important both for the theory of partition and for use as controls in subsequent empirical tests. Table 2 presents summary statistics for all variables and explains what each one measures. Table 3 presents a correlation matrix with the most important variables used in the analysis.

I now turn to the question of the determinants of partition.

MAIN DETERMINANTS OF PARTITION

My data set includes 125 civil wars, which produced 21 partitions.[44] Using the entire data set (which includes six right-truncated wars), I estimated probit models of the incidence of partition, selecting the explanatory variables on the basis of theory developed in the literature on civil war.[45] I want to test whether some of the same variables that either cause or terminate civil wars are also significant determinants of war-related partition.

I make the following testable hypotheses: Following the reasoning of partition theorists, I hypothesize that WARTYPE (ethnic/religious rather than ideological war) should be positively associated with parti-

[41] Kaufmann (fn.1, 1996 and 1998). Other studies also use the broad definition, given that the distinction between secession and partition seems artificial. See, among others, Horowitz (fn. 2); and Alexis Heraclides, *The Self-Determination of Minorities in International Politics* (Portland: Frank Cass, 1991); both use the terms partition and secession interchangeably.

[42] I consider only post–World War II cases because of the paucity of economic data from before 1945. Thus, I exclude the partition of Ireland. Cases of peaceful partition are also excluded, for example, Macedonia (1992), Czechoslovakia (1993), and Singapore (1965). I exclude one case (Iraq) that I believe was erroneously classified as a partition in Kaufmann (fn. 1, 1998). I exclude Iraq (1991) because there is no recognized, functional, or even autonomous Iraqi Kurdistan and the territory and its population would have been within reach of the Iraqi military had it not been for the U.S.-enforced no-fly zone.

[43] My coding of cases of partition incorporated suggestions made by anonymous referees.

[44] One might argue that the "real" number of partitions is smaller, since several of them occurred in either the former Yugoslavia or the former USSR. This would imply that these partitions may not be independent of one another. Thus, I cluster all same-country observations in my statistical analysis, relaxing the assumption of independence for those observations and allowing for nonconstant variance within clusters.

[45] See, for example, Paul Collier and Anke Hoeffler, "Justice-Seeking and Loot-Seeking in Civil War" (Manuscript, World Bank, February 1999); Michael W. Doyle and Nicholas Sambanis, "International Peacebuilding: A Theoretical and Quantitative Analysis" (Manuscript, Princeton University and the World Bank, February 2000); Paul Collier, "On the Economic Consequences of Civil War," *Oxford Economic Papers* 51 (1998); and Mason and Fett (fn. 33).

TABLE 1
CIVIL WARS BY YEAR AND TYPE, PARTITIONS, WAR RECURRENCE, AND LOW-LEVEL VIOLENCE

Country Name Where Civil War Took Place	*Year War Started*	*Year War Ended*	*Did War End for 2 years?*	*Did Residual Violence End for 2 years?*	*Was There a Partition?*	*Type of War (Identity or Not?)*
Afghanistan	1978	1992	no	no	no	ideology/other
Afghanistan	1993	ongoing	no	no	no	ethnic/religious
Algeria	1962	1963	yes	yes	no	ideology/other
Algeria	1992	1997	no	no	no	ethnic/religious
Angola	1975	1991	no	no	no	ethnic/religious
Angola	1992	ongoing	no	no	no	ethnic/religious
Argentina	1955	1955	yes	yes	no	ideology/other
Azerbaijan	1988	1996	yes	yes	yes	ethnic/religious
Bangladesh	1973	1994	yes	yes	no	ethnic/religious
Bolivia	1952	1952	yes	yes	no	ideology/other
Burma	1948	1951	yes	no	no	ideology/other
Burma	1968	1982	no	no	no	ethnic/religious
Burma	1983	1995	yes	no	no	ethnic/religious
Burundi	1965	1969	yes	no	no	ethnic/religious
Burundi	1972	1973	yes	yes	no	ethnic/religious
Burundi	1988	1988	no	no	no	ethnic/religious
Burundi	1991	ongoing	no	no	no	ethnic/religious
Cambodia	1970	1975	yes	no	no	ideology/other
Cambodia	1979	1991	yes	yes	no	ideology/other
Central African . Rep	1995	1997	yes	yes	no	ideology/other
Chad	1965	1979	no	no	no	ethnic/religious
Chad	1980	1994	yes	yes	no	ethnic/religious
China	1967	1968	yes	no	no	ethnic/religious
China-Taiwan	1947	1947	yes	no	yes	ideology/other
China-Tibet	1950	1951	yes	no	no	ethnic/religious
Colombia	1948	1962	yes	yes	no	ideology/other
Colombia	1978	ongoing	no	no	no	ideology/other
Congo Brazzaville	1992	1996	no	no	no	ideology/other
Congo/Zaire	1967	1967	yes	yes	no	ethnic/religious
Congo/Zaire	1975	1979	yes	no	no	ethnic/religious
Congo/Zaire	1960	1965	no	no	no	ethnic/religious
Congo/Zaire	1996	1997	no	no	no	ethnic/religious
Costa Rica	1948	1948	yes	yes	no	ideology/other
Cuba	1958	1959	yes	no	no	ideology/other
Cyprus	1963	1964	no	no	yes	ethnic/religious
Cyprus	1974	1974	yes	yes	yes	ethnic/religious
Djibouti	1991	1995	yes	yes	no	ideology/other
Dominican Rep.	1965	1965	yes	yes	no	ideology/other
El Salvador	1979	1992	yes	yes	no	ideology/other
Ethiopia/Eritrea	1974	1991	yes	yes	yes	ethnic/religious
Ethiopia	1977	1985	yes	no	no	ethnic/religious
Ethiopia	1974	1991	yes	yes	no	ideology/other

TABLE 1 *(cont.)*

Country Name Where Civil War Took Place	*Year War Started*	*Year War Ended*	*Did War End for 2 years?*	*Did Residual Violence End for 2 years?*	*Was There a Partition?*	*Type of War (Identity or Not?)*
Georgia	1991	1993	yes	yes	yes	ethnic/religious
Georgia	1992	1994	yes	yes	yes	ethnic/religious
Greece	1944	1949	yes	yes	no	ideology/other
Guatemala	1954	1954	yes	yes	no	ideology/other
Guatemala	1966	1972	no	no	no	ethnic/religious
Guatemala	1974	1994	yes	yes	no	ethnic/religious
Haiti	1991	1994	no	no	no	ideology/other
Haiti	1995	1996	yes	yes	no	ideology/other
India	1946	1948	yes	yes	yes	ethnic/religious
India	1965	1965	yes	no	yes	ethnic/religious
India	1984	1994	yes	yes	no	ethnic/religious
India	1989	1994	yes	no	yes	ethnic/religious
Indonesia	1956	1960	yes	no	no	ideology/other
Indonesia	1986	1986	yes	yes	no	ethnic/religious
Indonesia	1950	1950	no	no	no	ethnic/religious
Indonesia	1953	1953	no	no	no	ethnic/religious
Indonesia	1975	1982	yes	no	no	ethnic/religious
Iran	1978	1979	no	no	no	ideology/other
Iran	1981	1982	yes	no	no	ethnic/religious
Iraq	1959	1959	no	no	no	ethnic/religious
Iraq	1961	1975	yes	no	no	ethnic/religious
Iraq	1988	1994	yes	no	no	ethnic/religious
Iraq	1991	1994	yes	no	no	ethnic/religious
Israel/Palestine	1947	1949	no	no	yes	ethnic/religious
Israel	1950	1994	yes	yes	no	ethnic/religious
Jordan	1971	1971	yes	yes	no	ethnic/religious
Kenya	1991	1993	yes	no	no	ethnic/religious
Korea	1950	1953	yes	yes	yes	ideology/other
Laos	1960	1975	yes	no	no	ideology/other
Lebanon	1958	1958	yes	yes	no	ethnic/religious
Lebanon	1975	1978	yes	no	no	ethnic/religious
Lebanon	1982	1992	yes	no	no	ethnic/religious
Liberia	1989	1992	no	no	no	ideology/other
Liberia	1993	1996	yes	no	no	ideology/other
Malaysia	1948	1959	yes	yes	no	ideology/other
Mali	1990	1995	yes	yes	no	ethnic/religious
Mexico	1992	1994	yes	yes	no	ethnic/religious
Moldova	1992	1994	yes	no	yes	ethnic/religious
Morocco/ W. Sahara	1975	1989	yes	yes	no	ethnic/religious
Mozambique	1979	1992	yes	yes	no	ideology/other
Namibia	1965	1989	yes	yes	no	ethnic/religious
Nicaragua	1978	1979	no	no	no	ideology/other
Nicaragua	1981	1989	yes	yes	no	ideology/other

TABLE 1 *(cont.)*

Country Name Where Civil War Took Place	*Year War Started*	*Year War Ended*	*Did War End for 2 years?*	*Did Residual Violence End for 2 years?*	*Was There a Partition?*	*Type of War (Identity or Not?)*
Nigeria	1967	1970	yes	yes	no	ethnic/religious
Nigeria	1980	1984	yes	no	no	ethnic/religious
Northern Ireland	1968	1994	yes	yes	no	ethnic/religious
Pakistan	1971	1971	yes	yes	yes	ethnic/religious
Pakistan	1973	1977	yes	no	no	ethnic/religious
Papua New Guinea	1988	1991	yes	no	no	ethnic/religious
Paraguay	1947	1947	yes	yes	no	ethnic/religious
Peru	1980	1996	yes	no	no	ideology/other
Philippines	1950	1952	yes	yes	no	ideology/other
Philippines	1972	1996	yes	no	no	ethnic/religious
Philippines	1972	1992	yes	no	no	ideology/other
Romania	1989	1989	yes	yes	no	ideology/other
Russia/Chechnya	1994	1996	no	no	yes	ethnic/religious
Rwanda	1963	1964	yes	no	no	ethnic/religious
Rwanda	1990	1994	yes	yes	no	ethnic/religious
Sierra Leone	1991	1996	no	no	no	ideology/other
Somalia	1988	1991	no	no	yes	ethnic/religious
Somalia	1992	ongoing	no	no	no	ethnic/religious
South Africa	1976	1994	yes	yes	no	ethnic/religious
Sri Lanka	1971	1971	yes	yes	no	ideology/other
Sri Lanka	1987	1989	yes	yes	no	ideology/other
Sri Lanka	1983	ongoing	no	no	no	ethnic/religious
Sudan	1963	1972	yes	yes	no	ethnic/religious
Sudan	1983	ongoing	no	no	no	ethnic/religious
Tajikistan	1992	1994	yes	no	yes	ethnic/religious
Thailand	1967	1985	yes	yes	no	ethnic/religious
Turkey	1984	ongoing	no	no	no	ethnic/religious
Uganda	1966	1966	yes	yes	no	ethnic/religious
Uganda	1978	1979	no	no	no	ideology/other
Uganda	1980	1986	yes	no	no	ethnic/religious
Vietnam Rep.	1960	1975	yes	yes	yes	ideology/other
Yemen	1948	1948	yes	yes	no	ideology/other
Yemen	1994	1994	yes	yes	no	ideology/other
Yemen, North	1962	1969	yes	yes	no	ideology/other
Yemen, South	1986	1987	yes	yes	no	ideology/other
Yugoslavia/Bosnia	1992	1995	yes	yes	yes	ethnic/religious
Yugoslavia/Croatia	1991	1991	no	no	yes	ethnic/religious
Yugoslavia/Croatia	1995	1995	yes	yes	yes	ethnic/religious
Zimbabwe/ Rhodesia	1972	1980	yes	no	no	ethnic/religious
Zimbabwe	1984	1984	yes	yes	no	ethnic/religious

TABLE 2
SUMMARY STATISTICS OF VARIABLES USED IN THE ANALYSIS

Variable	*Proxy for*	*Obs.*	*Mean*	*Std. Dev.*	*Min*	*Max*
WAREND2	did the war end for 2 years?	125	0.744	0.438178	0	1
WAREND5	did the war end for 5 years?	114	0.745614	0.437438	0	1
NOVIOL2	2 years without low violence?	125	0.472	0.501224	0	1
NOVIOL5	5 years without low violence?	114	0.5	0.502208	0	1
PARTV2	is this a partitioned country?	125	0.168	0.375371	0	1
LOGCOST	log of deaths & displacements	124	11.93372	2.412293	6.907755	15.67181
LOGDEAD	log of deaths; battle & civilian	124	10.2669	2.093265	6.214608	14.91412
INTENSE	human cost per capita per month	124	0.005733	0.020525	4.83E-07	0.161202
OUTCOME2	how did the war end?	125	2.192	1.348213	0	4
TREATY	was a treaty signed?	125	0.28	0.450806	0	1
VREBEL	did the rebels win a victory?	125	0.216	0.41317	0	1
TRUCE	did the war end in a truce?	125	0.12	0.326269	0	1
MILOUT	was there a military outcome?	125	0.608	0.490161	0	1
WARDUR	duration of the war (months)	125	77.72	89.60987	1	528
GARM	size of gov't military (thousand)	124	402.729	957.5273	9	8256
ENERCAP	energy consumption per capita	124	0.000157	0.000413	−1.50e-07	0.00214
ELECTRIC	electricity consumption p/c (kwhrs)	123	586.8661	958.0384	10	5387
GDP	real GDP per capita (PPP)	124	1765.581	1789.364	130	10000
RGDPCH2	real GDP p.c. –no imputations	123	1556.398	1452.303	130	7741
PW10	was there a war in past 10 years?	125	0.48	0.50161	0	1
AIDGNP	foreign aid as % of GNP	116	8.731095	14.04966	0	83.8
ILLIT	percent illiterate population	118	43.23025	26.64874	1	85.5
GINI	income inequality index	125	40.3881	9.49447	19.29	62.88
LIFES	life expectancy at birth	125	52.15048	10.84891	32	73
EH	ethnic heterogeneity index	125	56.776	33.93563	0	144

TABLE 2 *(cont.)*

Variable	*Proxy for*	*Obs.*	*Mean*	*Std. Dev.*	*Min*	*Max*
EHLPOP	EH interacted w/ log of population	125	935.8059	594.8487	0	2629.118
ELF	ethnolinguistic fractionalization	118	48.66102	30.61764	0	93
LOGPOP	log of population size	125	16.27787	1.535166	13.26733	20.53999
GEO	dummy for continent	125	3.592	1.289511	1	5
DECADE	decade in which war started	125	3.44	1.433471	1	6
WARTYPE	was it an ethnic/identity war?	125	0.64	0.481932	0	1
TYPELICK	Licklider's version of wartype	116	1.293103	0.45716	1	2
TYPESTF	state failure project wartype	108	0.722222	0.449991	0	1
GURR2	democracy—2 years after war	109	9.311927	6.37515	1	20
IGURR2	Gurr2 w/ imputed missing cells	122	8.823376	6.310452	1	20
GPOL2	Freedom House political rights	88	4.840909	1.793259	1	7
GURRLAG5	5-yr mean democracy (prewar)	120	6.016667	5.680548	0	20
PEACEOP	third party peace operation type	125	1.056	1.477269	0	4
COLDWAR	dummy for the cold war period	125	.784	.4131703	0	1
PTEH	interaction term: EH*Partv2	125	9.6	26.57461	0	128
PTLCOST	interaction term: Logcost*Partv2	124	2.249478	5.031744	0	14.84513
PTARMY	interaction term: Garm*Partv2	124	140.6427	835.8922	0	8256
ETHPART	Interaction term: Wartype*Partv2	125	.144	.3525026	0	1

Table 3
Correlation Matrix of Key Variables (obs. = 99)

	WAREND2	WAREND5	NOVIOL2	NOVIOL5	PARTV2	LOGCOST	OUTCOME2	TREATY	VREBEL	TRUCE
WAREND2	1									
WAREND5	0.8315	1								
NOVIOL2	0.551	0.5831	1							
NOVIOL5	0.5136	0.595	0.9395	1						
PARTV2	0.0639	−0.0592	0.078	0.0702	1					
LOGCOST	−0.2017	−0.238	−0.211	−0.183	0.2743	1				
OUTCOME2	0.24[illegible]	0.2034	0.2394	0.2375	0.2151	0.3133	1			
TREATY	0.0982	0.1234	0.2115	0.1997	−0.0281	0.3194	0.792	1		
VREBEL	0.0311	0.0575	−0.0874	−0.0527	0.1182	0.0104	−0.0591	−0.3469	1	
TRUCE	0	−0.1491	−0.0447	−0.0512	0.3972	0.0897	0.2126	−0.1887	−0.1838	1
MILOUT	−0.1308	−0.0617	−0.1909	−0.1764	−0.2227	−0.3245	−0.8978	−0.7768	0.4204	−0.4372
WARDUR	−0.0716	−0.0527	0.0394	0.0294	−0.1496	0.2662	0.2539	0.3623	−0.0807	−0.0859
GARM	0.1439	0.1139	−0.0655	−0.0736	0.2805	0.0658	−0.0682	−0.1726	−0.0372	0.2271
ENERCAP	0.1244	0.1206	0.2102	0.2033	0.2496	−0.0298	0.3037	0.2274	−0.1076	0.1899
GDP	0.0704	0.0963	0.1282	0.1147	0.0431	−0.1695	0.1574	0.1216	−0.0586	0.0899
EH	−0.0763	−0.1779	−0.0907	−0.156	0.1197	0.1456	0.0876	0.0899	−0.0022	−0.0051
EHLPOP	−0.0504	−0.1812	−0.102	−0.1631	0.1619	0.1371	0.0474	0.0378	−0.0195	0.0341
ELF	−0.0293	−0.0598	−0.0558	−0.1404	0.0767	−0.0178	0.0611	0.0744	−0.0843	0.0487
LOGPOP	0.0783	−0.044	−0.1768	−0.2017	0.2106	0.0433	−0.2587	−0.2654	−0.0579	0.0705
GEO	−0.0513	−0.0645	−0.1747	−0.2006	−0.1156	0.1143	−0.1046	−0.0138	−0.0949	−0.0405
DECADE	−0.191	−0.1595	−0.1097	−0.0985	0.0187	0.1561	0.0913	0.1472	−0.1039	0.0989
WARTYPE	−0.101	−0.1826	−0.1872	−0.2141	0.1074	0.0563	−0.0369	0.0694	−0.3339	0.093
PW10	0.0054	−0.0414	−0.1923	−0.2129	0.1456	0.1777	−0.0558	−0.006	−0.0291	0.0319
GURR2	0.1874	0.144	0.3281	0.358	0.1224	−0.0584	0.286	0.2708	−0.2439	0.18
GURRLAG5	0.0884	−0.0286	0.0744	0.036	−0.0184	−0.2111	0.0337	0.0341	−0.2133	0.1562

TABLE 3 *(cont.)*

	WAREND2	*WAREND5*	*NOVIOL2*	*NOVIOL5*	*PARTV2*	*LOGCOST*	*OUTCOME2*	*TREATY*	*VREBEL*	*TRUCE*
PEACEOP	0.0412	0.0733	0.1117	0.1255	0.2226	0.3622	0.4809	0.4213	−0.1693	0.3123
PTEH	0.0635	−0.1367	0.0285	0.0221	0.822	0.2579	0.2187	0.0405	0.037	0.32
PTLCOST	0.0648	−0.0545	0.0964	0.0886	0.9943	0.2937	0.218	−0.0182	0.1153	0.3833
PTARMY	0.0906	0.0422	0.0329	0.0294	0.4481	0.0962	0.0754	−0.0965	0.0302	0.3408
ETHPART	0.0179	−0.1233	0.0569	0.0502	0.8621	0.2541	0.2254	0.0285	0.0366	0.3604

	MILOUT	*WARDUR*	*GARM*	*ENERCAP*	*GDP*	*EH*	*EHLPOP*	*ELF*	*LOGPOP*	*GEO*
MILOUT	1									
WARDUR	−0.2824	1								
GARM	0.0131	−0.09	1							
ENERCAP	−0.3326	0.0801	−0.0723	1						
GDP	−0.1791	0.2231	−0.0532	0.5168	1					
EH	−0.076	0.1605	−0.153	−0.0741	−0.1648	1				
EHLPOP	−0.0529	0.1331	−0.0993	−0.1239	−0.1813	0.983	1			
ELF	−0.0776	0.1788	−0.1523	−0.0739	−0.067	0.6744	0.6919	1		
LOGPOP	0.1985	−0.0425	0.4709	−0.4209	−0.1725	0.219	0.3585	0.3016	1	
GEO	0.0608	0.1599	0.0757	−0.446	−0.576	0.2956	0.313	0.3914	0.2669	1
DECADE	−0.1687	0.0442	−0.2195	0.0152	0.1828	0.0667	0.0643	0.0425	−0.0308	0.0381
WARTYPE	−0.1045	0.0594	−0.0225	0.1301	0.0432	0.3144	0.3221	0.3169	0.1356	0.3131
PW10	0.0073	0.1091	0.2128	−0.044	−0.1404	0.1435	0.1533	0.0553	0.2389	0.2513
IGURR2	−0.371	0.2454	−0.1826	0.2337	0.2378	−0.0007	0.0274	0.0583	−0.0277	−0.2358
GURRLAG5	−0.1377	0.0949	−0.047	0.1116	0.127	0.1147	0.1447	0.1075	0.0415	−0.0871
PEACEOP	−0.5503	0.0866	0.0266	0.2812	0.1221	0.0195	−0.0141	−0.008	−0.2631	−0.1399
PTEH	−0.2381	−0.1141	0.0102	0.1687	−0.0415	0.3518	0.4155	0.2268	0.28	−0.053

TABLE 3 *(cont.)*

	WAREND2	*WAREND5*	*NOVIOL2*	*NOVIOL5*	*PARTV2*	*LOGCOST*	*OUTCOME2*	*TREATY*	*VREBEL*	*TRUCE*
PTLCOST	−0.2234	−0.1415	0.266	0.2412	0.0353	0.1329	0.1742	0.0825	0.2018	−0.1079
PTARMY	−0.1206	−0.0929	0.8294	−0.0539	−0.1074	−0.1418	−0.1124	−0.1571	0.2031	0.0574
ETHPART	−0.2517	−0.1618	−0.0377	0.3132	0.0773	0.2611	0.2989	0.2046	0.1296	−0.161

	DECADE	*WARTYPE*	*PW10*	*IGURR2*	*GURRLAG5*	*PEACEOP*	*PTEH*	*PTLCOST*	*PTARMY*	*ETHPART*
DECADE	1									
WARTYPE	0.228	1								
PW10	0.2961	0.2596	1							
IGURR2	−0.033	−0.1643	−0.2346	1						
GURRLAG5	−0.1262	−0.0525	−0.165	0.5018	1					
PEACEOP	0.2453	0.0979	0.0839	0.2047	−0.0326	1				
PTEH	0.0903	0.2149	0.0621	0.2121	0.0784	0.1812	1			
PTLCOST	0.0163	0.1128	0.1342	0.1241	−0.0233	0.2248	0.835	1		
PTARMY	−0.2142	−0.1507	0.1349	−0.0866	−0.0511	0.1655	0.1181	0.4309	1	
ETHPART	0.1625	0.2534	0.0637	0.2335	0.0113	0.2277	0.9107	0.8668	0.0689	1

:ion.[46] Partition is not usually the goal of revolutions motivated by ide- ɔlogy, but it is often the stated goal of ethnic war. For the same reason, :thnic heterogeneity (EH) should be significantly associated with parti- ion, especially if ethnic groups are large enough to constitute a politi- :ally and economically viable successor state. The human toll of the war 'LOGCOST) should be positively correlated with partition. The intuition s that extremely violent wars can be settled only by partition and that he international community might be more supportive of such an out- :ome in those cases. The outcome of the war (OUTCOME2) should also ɔe a significant determinant of partition, since we would not expect to see partition if the government wins a military victory, whereas parti- ion would be more likely in the case of rebel victory. Population size LOGPOP) should be positively associated with partition.[47] I also include ıs control variables a number of socioeconomic indicators of the coun- ry's overall level of development, since many studies have identified uch variables as significant determinants of civil war.[48] However, it is ıarder to theorize about the nature of their association with the inci- lence of partition. I present the results of my estimations in Table 4.

A first important finding is that, as theorized, the type of the war is ndeed a significant determinant of partition. Identity wars (ethnic and eligious wars) are positively and significantly correlated with partition n model 1. This implies that the partition theorists correctly argue that thnicity matters for the onset of partition. However, when I tested al- ernatively coded variables for the type of war in models 1b and 1c, I ound a significant difference: both variables were significant, but Lick- ider's variable (TYPELICK in model 1a) is negatively correlated with the ɔnset of partition whereas the State Failure Project variable (TYPESTF n model 1b) is positively correlated with partition. This implies that, vhile ethnicity seems important for partition, the direction of their as- ociation might be influenced by assumptions made in the coding of he WARTYPE variable.[49] At the same time, only small changes are ob-

[46] The coding of the WARTYPE variable was not easy. There are substantial differences in the various ›urces and data sets I consulted. I used two main sources for this variable: Licklider (fn. 32); and Esty t al. (fn. 33). I coded the variables TYPELICK (Licklider's war-issue variable) and TYPESTF (the State- ailure Project's war-type variable) to facilitate comparisons across cases. Where those two sources dif- :red, I coded WARTYPE based on majority opinion in other data sets, including Regan (fn. 33); and ⁄lason and Fett (fn. 33).

[47] This hypothesis (with an emphasis on the proportion of young men) has been posited with refer- nce to the causes of civil war by Collier and Hoeffler (fn. 45); and Robert H. Bates, "Ethnicity, Cap- al Formation, and Conflict," *CID Working Paper* no. 27 (Harvard University, October 1999).

[48] For example, Collier and Hoeffler (fn. 45).

[49] Given such problems, quantitative analysts of civil wars must be highly transparent in their coding f these variables. Moreover, it is necessary for those building data sets of civil war to coordinate their

TABLE 4
PROBIT REGRESSION OF OCCURRENCE OF PARTITION[a]

Dep. Var.: Partition	*Model 1*	*Model 1a*	*Model 1b*	*Model 2*	*Model 3*
Constant	−4.53***	−2.09	−3.93*	−6.72*	−2.38
	(1.64)	(1.82)	(2.12)	(3.48)	(2.09)
WARTYPE	1.65***			2.22***	1.59***
(identity or	(.553)	—	—	(.822)	(.547)
ideology?)	.103			2.22	.032
LOGCOST	.358***	.471***	.338**	.441***	.396**
(log of deaths &	(.118)	(.141)	(.138)	(.152)	(.138)
displacements	.025	.025	.034	.44	.008
EH	−.064**	−.068**	−.063**	−.049	−.075**
(Index of Ethnic	(.026)	(.029)	(.027)	(.035)	(.027)
Heterogeneity)	-.005	-.004	-.006	−.049	−.0015
EHLPOP	.003**	.0037**	.0034**	.003	.004**
(EH * log of	(.001)	(.002)	(.0015)	(.002)	(.0014)
population size)	.0002	.0002	.0003	.003	.0000[illegible]
TRUCE	1.69***	1.76***	1.71***	1.58***	1.75***
(war ended in an	(.348)	(.383)	(.342)	(.438)	(.414)
informal truce?)	.337	.313	.401	1.58	.183
VREBEL	1.55***	1.59***	1.35***	1.91**	2.16***
(war ended with a	(.519)	(.578)	(.425)	(.778)	(.502)
rebel victory?)	.247	.214	.256	1.91	.229
GEO	−.594***	−.644***	−.651***	−.514**	−.405*
(continent)	(.172)	(.205)	(.192)	(.207)	(.225)
	−.042	−.035	-.065	−.513	−.008
DECADE	−.175	−.204	−.167	−.019	−.023
(decade in which	(.140)	(.161)	(.151)	(.181)	(.153)
war started)	−.012	−.011	−.017	−.019	−.0005
TYPELICK		−2.07***			
(Licklider's wartype)		(.591)			
		−.113			
TYPESTF (State			1.32**		
Failure			(.576)		
Project wartype)			.094		
ILLIT (percentage				−.021	
population				(.016)	
that is illiterate)				−.02	
GDP (real GDP				2.17e-07	
per capita, PPP)				(.0001)	
				2.17e-07	
LIFES (life				−.002	
expectancy at birth)				(.039)	
				−.002	
GINI (income					−.113***
inequality index)					(.029)
					−.002
ENERGYS (energy					.0003
consumption					(.0002)
per capita)					5.72e-06

E 4 *(cont.)*

ar.: Partition	*Model 1*	*Model 1a*	*Model 1b*	*Model 2*	*Model 3*
vations	124	115	107	116	123
kelihood	−30.365	−27.397	−29.242	−22.042	−21.529
)-R^2	0.4616	0.4844	0.4481	0.5597	0.6170
:tly classified	87.20%	87.82%	86.91%	92.24%	93.49%
tion in error	23.80%	27.55%	22.08%	53.80%	61.25%

ignificant at the .01 level; ** significant at the .05 level; * significant at the .10 level (two-tailed tests)
ported are coefficients (robust standards errors) and marginals (dF/dx), in that order.

rved in most of the other variables in models 1a and 1b, which sug-
:sts that the other variables are robust.

Ethnicity has a complicated relationship with partition. In models 1, ι, and 1b, I found that as ethnic heterogeneity increases, the probabil-/ of a partition decreases significantly, suggesting that it may be diffi-ılt to coordinate and win in a secessionist war in extremely diverse cieties.[50] However, as the size of ethnic groups increases, so does the obability of partition.[51] Large ethnic groups may be better able to ercome the coordination problems associated with mounting a rebel-ın and better able to defend their territory.

As expected, I found that partitions are positively and significantly rrelated with the level of violence (LOGCOST). This variable is very bust, and since deaths and displacements chronologically precede the currence of partition, these results may be pointing to a causal rela-ınship between high levels of violence and partition. At the same ne, my data do not allow me to preclude the possibility that some of e observed violence may actually have been caused by partition (as, r example, in India and Cyprus).

War outcomes are also significant in model 1 and all its variants. ıth rebel victory and truce are significantly and positively associated th the incidence of partition (although, these results should be inter-eted with caution due to high collinearity among these regressors).

In model 2 I added several socioeconomic variables that may have en important causes of the previous war and I found them individu-y insignificant. While I did not expect to find individual significance

ırts and exchange information. I will use my WARTYPE variable in the rest of the analysis since I gen-ted it with reference to as many sources as I could consult for each case. I tried to reflect majority nion about the coding of each case, where there was disagreement between my main sources.

[0] If we drop WARTYPE, then EH becomes nonsignificant but remains negative.

[1] I proxy the size of ethnic groups by interacting the ethnic heterogeneity index EH with the log of ıulation size (LOGPOP).

given their high level of collinearity, I did expect and did find joint significance (a joint test of ILLIT, LIFES, GDP, DECADE, and GEO yielded chi2(5) = 11.85 and Prob > chi2 = 0.0369). The fact that these variables are not individually significant allows me to use them as controls in my analytical models of war recurrence in the next section, since they are not significantly correlated with partition and I use partition as a core regressor in those models. The most important impact of these local capacity variables is that they make ethnicity (EH, EHLPOP) nonsignificant (although this may be due to the noise that they introduce in the equation).

Finally, in model 3 I controlled for income inequality (GINI) and replaced the socioeconomic controls of model 2 with a variable measuring the country's overall level of economic development (ENERGYS) at the start of the war.[52] Here I found a positive but nonsignificant relationship between economic development and partition. The opposite association is often found between development and the risk of onset of war, but my finding makes sense, since low levels of economic development often discourage ethnic minorities from seceding (although a more accurate result might have been obtained if I had data available on the regional concentration of natural resources and the geographic dispersion of industries within each country).[53] Income inequality is significant but negatively correlated with partition, which once again seems counterintuitive.[54] The control variables for geographical location and for the decade during which the war started pick up time- and place-specific effects. The DECADE variable is nonsignificant, but there seem to be important regional effects (which could be better studied in the context of a panel data set).

To summarize, I found that partition is significantly more likely to occur after an identity war than after an ideological war, after an informal truce or rebel victory following a very costly war, in a country with

[52] Measuring these variables at the start of the war not only prevents problems of reverse causality but also captures any impact that these variables might have had on causing the civil war in the first place. Thus, the inability to find significance for the economic variables in model 2 may be due to a selection effect (since all the countries in my sample are countries that experienced war and may therefore share the same socioeconomic background). Thus, the analysis of partition and war recurrence must focus here on war-related variables that would be expected to differ significantly across countries.

[53] See Collier and Hoeffler (fn. 45). The precise relationship between partition and economic variables is undertheorized, so I will not explore this further, but this counterintuitive finding is worth further study. It may be that relatively richer countries can support partition, since the prospects of economic viability of the successor state will be greater.

[54] These signs do not change if I drop the cases of ongoing war. The direction of this correlation, however, may result from measurement error or selection effects. Measurement error is possible because reliable data were often not available for the relevant years. Or it may be due to collinearity between income inequality and energy consumption, since I used GDP data to impute missing values of both of these variables.

large ethnic groups and small levels of ethnic heterogeneity and a relatively higher level of economic development (compared to other war-torn states). Having identified these correlates of partition, I can now test the three critical hypotheses of partition theory: (1) that partitions create successor states that are at least as democratic as their predecessors, if not more so; (2) that partitions reduce the risk of war recurrence; and (3) that partitions reduce low-level ethnic violence after the war ends.

TESTING THE HYPOTHESES

DOES PARTITION CREATE UNDEMOCRATIC STATES?

Kaufmann has argued that successor states are generally no less democratic than their predecessors and that they can even be more democratic.[55] Although a full test of this hypothesis is not yet possible (see below), a first cut at such a test is provided in this section.

Using data on democracy and autocracy from the Polity 98 data set, I have created the variable GURR, measuring the level of democracy two years (GURR 2) and five years (GURR 5) after the end of the war for both predecessor and successor states.[56] Partitioned countries have a mean and standard deviation of GURR 2 of 11.02 and 6.204, respectively. The corresponding values for nonpartitioned countries are 8.42 and 6.27, respectively. A score of 20 suggests a perfect democracy, whereas a score of 0, an extreme autocracy. The computed averages reveal that countries that have experienced a civil war—regardless of whether or not they have been partitioned—are generally nondemocratic two years after the end of the war. This may be due to the war itself or to a legacy of undemocratic institutions. These legacies can be measured by GURRLAG5—the mean level of democracy during the five years prior to the start of the war. Table 5 lists all partitioned countries and their GURR and GURRLAG5 indices. Thirteen cases broadly support Kaufmann's hypothesis that partitioned countries are no less democratic than their predecessors and six cases do not support that hypothesis.[57]

[55] Kaufmann (fn. 1, 1998), 124.

[56] The original data on democracy were compiled by Keith Jaggers and Ted Robert Gurr, *Polity 98 Project (http://www.bsos.umd.edu/cidcm/polity/)*. I added their democracy and autocracy scores as follows: GURR = [DEMOCRACY+ (10-AUTOCRACY)]. The resulting variable ranges from 0 to 20. The Polity3 data end in 1994, so I imputed thirty-five missing values using the political rights index of the Freedom House project after I established that there was a very close correlation between Gurr's democracy index and Freedom House's political rights index. See Freedom House, *Freedom in the World* (London: Freedom House, 1999).

[57] This list includes not only ethnic partitions but also all other cases of partition in my data set. Subsequent analysis focuses directly on partitions that resulted from ethnic wars and therefore excludes a number of partitions (for example, the Koreas, Vietnam, and Taiwan). However, I test the robustness of my results by including all wars and partitions.

TABLE 5
PREWAR AND POSTPARTITION POLITY INDICES[a]

Country Name	*Five-Year Prewar Democracy Index (GURRLAG5)*	*Democracy Index Two Years after the War (GURR2)*
Azerbaijan	3	4
Yugoslavia-Bosnia	6	9
China-Taiwan	5	2
Yugoslavia-Croatia	6	11
Yugoslavia-Croatia	6	9
Cyprus	1	—
Cyprus	1	20
Ethiopia-Eritrea	1	11
Georgia-Abkhazia	11	15
Georgia-Ossetia	11	15
India-Pakistan	1	16
India-Kashmir	19	17
India-Kashmir	18	18
Israel-Palestine	—	—
Korea (North-South)	1	3
Moldova	1	17
Pakistan-Bangladesh	3	18
Russia-Chechnya	11	14
Somalia-Somaliland	3	2*
Tajikistan	11	5
Vietnam, Republic of	7	3

[a] GURR2 includes one imputed value (Somalia), denoted with an asterisk and rounded to tl nearest integer. If the predecessor state was a colony or not a recognized state (e.g., Israel) 1–5 yea before the war started, then we do not have a Polity score since these are only available for independe states. In those cases, I have entered an N/A and comparison of the polity index before and after tl war is not possible. The same is true for cases where the country was created out of the civil war, as Israel. A dot indicates no available information. The index is based on data from Polity 98 (see text)

An equality of means test for both GURRLAG5 and GURR2 revea signs of significant differences between partitioned and nonpartitione countries. Specifically, a t-test of the null hypothesis that there is n significant difference in the mean of GURR2 in partitioned and non partitioned countries can marginally be rejected with 120 degrees c freedom and a t-statistic of −1.6644 ($P > |t| = 0.098$). A two-sampl t-test with equal variances only on cases of ethnic war does reveal th there is a significant difference and that partitioned countries hav higher democracy averages (75 d.f.; $t = -3.3842$). However, this effe need not refer to the democratization effect of partition and may be du to the prewar level of democracy, which is on average higher amon

partitioned countries that have experienced war. A means test with respect to DEMCH—a variable that measures the difference of postwar and prewar levels of democracy—rejects equality with 73 degrees of freedom and $t = -2.6597$ ($P > |t| = 0.0096$). These results suggest that partitioned states fare slightly better in terms of postwar democratization.

I turn next to a multivariable OLS regression of GURR2 (and GURR5) using partition as one of the explanatory variables. A bivariate regression of democracy on partition shows no significance of partition at the 5 percent level if we use the entire data set but finds partition positive and significant among cases of ethnic war. A clearer picture emerges from multivariate models (see Table 6). In model 1 I regressed the postwar democracy index (with imputed missing values) on a number of explanatory variables, including partition.[58] Most of the variance in the dependent variable in this and the other models is explained by the country's democratic legacy (GURRLAG5), which is extremely robust and significantly increases postwar levels of democracy. There is a weak positive correlation between democratic postconflict states and the presence of third-party peace operations (PEACEOP). Given that this association is weak, I dropped PEACEOP from subsequent regressions. There is a negative, though not very robust relationship between the level of democracy and the size of the government's military, which suggests that troop demobilization and force reductions in postconflict states may be useful in promoting peace and democracy. War duration is significant and positively correlated with the level of democracy, lending support to the war-weariness hypothesis (that people tired of war will try harder to build peace). Real per capita GDP is positively correlated with democracy (as would be expected), but this association is not significant, possibly due to selection effects or measurement error.[59] Partition is positively correlated with democracy and it is significant at the 10 percent level. The type of war is also significant and the coefficient sign suggests that ethnic wars would reduce the postwar mean of democracy by 2.5 points as compared with nonethnic wars.

In model 2 I looked more closely at cases of ethnic war, and the results discussed previously are nearly identical, except for the signifi-

[58] I selected these variables based on theoretical arguments regarding the determinants of the level of democratization after civil war, drawing on Doyle and Sambanis (fn. 45), among others. Also the relationship between economic variables and democracy has been the focus of numerous studies in the economics and political science literatures; see, e.g., Ross E. Burkhart and Michael S. Lewis-Beck, "Comparative Democracy: The Economic Development Thesis," *American Political Science Review* 88 (December 1994).

[59] These regression results are robust for a large number of specifications.

Table 6
OLS and 2SLS Regressions of Postwar Democracy w/ Robust Standard Errors and Clustered Same-Country Observations[a]

Dependent Variable: Democracy 2 Years after End of the War (GURR2)	*Model 1* *All wars*	*Model 2* *Only Ethnic Wars*	*Model 3* *Nonimputed* GURR2	*Model 4*[b] *2SLS* *Only Ethnic*	*Model 5*[c] *2SLS* *Only Ethnic*	*Model 6*[d] *2SLS* *Only Ethnic*
Constant	4.56***	2.07	2.63*	1.46	1.11	1.57
	(1.42)	(1.26)	(1.33)	(1.33)	(1.20)	(1.18)
PARTITION (did the war	3.33*	5.72***	6.26***	7.53***	8.01***	6.01***
result in partition?)	(1.82)	(1.54)	(1.67)	(2.62)	(2.56)	(2.05)
WARTYPE (was it an	−2.42**					
identity or nonidentity war?)	(1.13)					
PEACEOP (were there UN	.66*					
or other peace operations?)	(.36)					
WARDUR	.015**	.02**	.023***	.023***	.024***	.021***
(duration of war in months)	(.006)	(.007)	(.008)	(.008)	(.008)	(.007)
GARM * 1,000 (size of the gov-	−1.12***	−.95*	−1.1*	−.77		
ernment military (in thousands))	(.27)	(.56)	(.6)	(.58)		
GDP * 1,000 (real income per	.5695	.53	.34	.46	.43	.51
per capita, PPP-adjusted)	(.45)	(.54)	(.56)	(.55)	(.55)	(.55)
GURRLAG5 (5-yr average	.48***	.47***	.43***	.46***	.45***	.46***
prewar democracy score)	(.08)	(.102)	(.11)	(.11)	(.11)	(.11)
Observations	116	74	64	73	74	74
Goodness-of-fit	$F(7, 61) = 18.47$	$F(5, 38) = 12.24$	$F(5, 33) = 10.38$	$F(5, 38) = 10.22$	$F(4, 39) = 10.87$	$F(4, 39) = 4.80$
	Prob > F = 0.0000	Prob > F = 0.0000	Prob > F = 0.0000	Prob > F = 0.0000	Prob > F = 0.0000	Prob > F = 0.0000
R^2	0.4180	0.5010	0.5069	0.4924	0.4794	0.4963

*** significant at the .01 level; ** significant at the .05 level; * significant at the .10 level

[a] Coefficients and standards errors are reported in parentheses.

[b] Instrumented: partition; instruments: LOGCOST OUTCOME2 EH EHLPOP GEO.

[c] Instrumented: partition; instruments: LOGDEAD OUTCOME2 EH EHLPOP GEO.

[d] Instrumented: partition; instruments: LOGCOST TRUCE VREBEL EH EHLPOP GEO.

cance and coefficient of partition, which now more than doubles. In model 3 I regressed the same right-hand-side variables on the GURR2 variable without imputing missing values, and the results are robust. The democratic legacy variable (GURRLAG5) seems to be doing heavy lifting in these regressions. If we drop it from models 2–3, the R^2 drops by about 20 percentage points and the coefficient of partition increases by 20 percent.

Partition may also be endogenous or, rather, jointly determined with some of the other right-hand-side variables. Thus, I reestimated model 2 using a two-stage least squares (2SLS) estimator; no significant differences were observed (both the coefficient and standard error of partition were extremely volatile, but the t-test remained the same). In model 5 I estimated a different specification of this model again using 2SLS, dropping the now nonsignificant GARM variables, but again no major change occurred. In model 6 I changed two of the instrumental variables used in the first-stage regressions. Throughout these changes in specification and estimation method, the democratic legacy variable (GURRLAG5) was extremely robust, as was partition. The coefficient of the partition variable was positive and varied between 6 and 8.

Finally, testing the hypothesis that the instrumented 2SLS models might be more efficient than the OLS models, I found the OLS models more efficient. A Hausman test of the null hypothesis that the difference in coefficients between the instrumented and OLS models is not systematic, estimated as: $(b - B)'[(V_b - V_B)^{-1}](b - B)$, yielded a chi-square statistic of 0.57 with 4 degrees of freedom, so we cannot reject the null hypothesis that the instrumented models are less efficient and we should rely on the OLS estimates of models 1–3.[60]

Overall, these regression results seem to support partition theory. However, this is at best an incomplete picture, since we do not have comparable democracy data for successor states that are not internationally recognized. This therefore creates a problem of systematic bias in the coding of the dependent variable, which reduces the reliability of the findings just described. Successor states that are not internationally recognized may well have low democracy; contrast, for example, the political institutions of the Republic of Cyprus with those of the so-called Turkish Republic of Northern Cyprus or the democracy levels in Chechnya versus those in Russia or in Ossetia and Abkhazia versus

[60] J. Hausman, "Specification Tests in Econometrics," *Econometrica* 46 (1978), 1251–71. On the use of the Hausman test to test exogeneity, see B. H. Baltagi, *Econometrics* (New York: Springer-Verlag, 1998), 291; and Stata Corporation, *Stata Reference Manual: Release 6* (College Station, Tex.: Stata, 1999), 2:7–13. In the formula above b is the coefficient vector from the consistent estimator and B the coefficient vector from the efficient estimator and V_b, and V_B are their respective covariance matrices.

those in Georgia. These comparisons are difficult, given that the Polity 98 project does not provide information on these territories. Thus, before making a judgment on the democratizing effects of partition, we need comparable data on democracy for all successor and predecessor states.[61] At this stage, I can neither reject nor accept the claims of the partition theorists with confidence. The link between partition and democratization must be studied further as high-quality data become available. At that time it would also be interesting to see whether high levels of GURRLAG5 tend to promote democracy in postwar successor states. At this stage the robustness of GURRLAG5 suggests that our focus should not be on designing democracy-friendly partitions but rather should be on strengthening democratic institutions in countries before they actually experience a civil war.

I now turn to the critical question of war recurrence.

Does Partition Prevent War Recurrence?

Several examples of wars following partitions provide support for the critics of partition theory: Croatia fought a second war with Serbia after it was partitioned in 1991. Ethiopia and Eritrea fought a bitter territorial war in 1999–2000 after being partitioned in 1991. The partition of Somaliland collapsed in a wave of new violence in 1992. India and Pakistan have fought three wars since their partition in 1947. Cyprus was at war again in 1974 after it was effectively partitioned into militarily defensible, self-administered enclaves between 1963 and 1967.[62] At the same time, seemingly intractable conflicts and bloody ethnic wars have given way to peace without partition, as, for example in South Africa, Guatemala, and Uganda.

[61] Given the paucity of data to answer this important question, a worthwhile project would be to conduct a comparative case study of the political institutions of all these successor states.

[62] This is not a well-known case. In 1963 a "green line" was established in the capital city of Nicosia partitioning the Greek Cypriot and Turkish Cypriot sectors. After 1964 the partition was expanded and more than 30 percent of the Turkish Cypriot population moved to defensible, self-administered enclaves. These enclaves forcibly excluded the Greek Cypriot population and their demilitarization and defortification was part of the mandate given to a UN peacekeeping force—UNFICYP (see UN doc S/5764, 15 June 1964, para. 61). The UN secretary-general often noted in his report that the enclaves gave the Turkish Cypriots "complete military and administrative control" of several areas (S/6228, para 50). In 1965 the secretary-general noted that the enclave fortifications "contribute to maintaining tension at high pitch" and UNFICYP "insists on their removal" (S/6228, para. 51). Within six months in 1967, 52 new positions were built by the Greek Cypriot National Guard and 130 by the Turkish Cypriots (S/8286, December 8, 1967, para. 50). The secretary-general noted that "this ceaseless building of fortifications . . . [would] result in the Island being criss-crossed and honeycombed with defences [*sic*]" (S/8286, para. 49). Thus, the island was effectively partitioned between 1963 and 1967. On the Cyprus conflict during the critical years between 1963 and 1974, see Richard Patrick, *Political Geography and the Cyprus Conflict, 1963–1971* (Waterloo, Canada: Department of Geography, University of Waterloo, 1976); the work includes maps of the pre-1974 enclaves. See also Joseph Joseph, *Cyprus: Ethnic Conflict and International Solution* (New York: St. Martin's, 1997); and Sambanis (fn. 33).

TABLE 7
CROSS-TABS BETWEEN PARTITION, WAR RECURRENCE, AND RESIDUAL VIOLENCE
(TWO, FIVE, AND TEN YEARS AFTER THE WAR)

War Ended for	*War Recurrence*	*Partitition*	*No Partitition*	*Total Cases*	*Pearson Chi-Square Test (1 Degree of Freedom)*
2 years	war recurred	5	27	32	Pearson chi2(1) =
	no war	16	77	93	0.0425; Pr = 0.837
5 years	war recurred	5	24	29	Pearson chi2(1) =
	no war	14	71	85	0.0092; Pr = 0.923
10 years	war recurred	5	24	29	Pearson chi2(1) =
	no war	6	46	52	0.5159; Pr = 0.473

War Ended for	*Residual Violence*	*Partitition*	*No Partitition*	*Total Cases*	*Pearson Chi-Square Test (1 Degree of Freedom)*
2 years	low violence	10	56	66	Pearson chi2(1) =
	no violence	11	48	59	0.2719; Pr = 0.602
5 years	low violence	9	48	57	Pearson chi2(1) =
	no violence	10	47	57	0.0632; Pr = 0.802
10 years	low violence	8	38	46	Pearson chi2(1) =
	no violence	5	32	37	0.2334; Pr = 0.629

These examples, however, do not constitute sufficient proof that the critics are right. To develop better insight into the relationship between partition and war recurrence, I present simple cross-tabulations of partition and war recurrence and low-level violence in Table 7.[63] No statistical association between partition and ending violence is evident either for the entire population of cases or for ethnic wars.

A fuller test of the relationship between partition and war recurrence, conditional on the effects of several other variables, can be derived by estimating multivariable regressions (see Table 8). The dependent variable is WAREND (did the war end?) with the suffixes 2, 5, and 10 denoting that it is observed two, five, and ten years after the end of the civil war. WAREND is coded 1 if there is no war recurrence and 0 otherwise.[64] The control variables in these regressions were selected on

[63] This table includes cases of nonethnic partition. I also ran these cross-tabs excluding nonidentity (ethnic/religious) wars, and the results were not significantly affected.

[64] I estimate probit models with clustered same-country observations and robust standard errors. Since partition theory has focused on ethnic wars and since I found the type of war to be a significant determinant of partition, I dropped cases of nonidentity wars from my analysis, but I do report some results of interest as they apply to all civil wars.

TABLE 8
MODELS OF CIVIL WAR RECURRENCE[a]

Dependent Variable: WAREND2 (Did the War End for Two Years?)	Model 1 Only Ethnic Wars Assume Exogeneity	Model 2 All Wars Assume Exogeneity	Model 3[b] 2SLS Linear Prob. Assume Endogeneity	Model 4[c] 2S Probit w/ Rivers & Vuong Test	Model 5[d] Bivariate Probit rho-Test for Exogeneity	Model 6 Probit w/ Partition Interaction Only Ethnic Wars	Model 7 Same as Model 6, but for the 5-Year Period	Model 8 Probit w/ Partition Interaction All Wars
PARTITION	−.068	.253	−.025	−.080	−.260	−4.72	−2.25	
did the war result	(.461)	(.365)	(.204)	(.694)	(.675)	(3.82)	(3.83)	—
in partition?	−.021	.066	—	—	—	−.962	−.739	
RES—residual to				.122				
test for	—	—	—	(.956)	—	—	—	—
exogeneity								
LOGCOST	−.164*	−.185***	−.049*	−.152*	−.146*	−.162*	−.185*	−.17***
natural log of	(.087)	(.067)	(.026)	(.086)	(.086)	(.096)	(.097)	(.07)
dead & displaced	−.051	−.052	—	—	—	−.049	−.058	−.050
OUTCOME2	.53***	.37***	.131***	.406***	.418***	.578***	.522***	.35***
war outcome—	(.147)	(.112)	(.036)	(.147)	(.142)	(.165)	(.145)	(.116)
four outcomes	.165	.104	—	—	—	.178	.166	.097
WARDUR	.0015	.0028*				.0006	.0016	.002
duration of war	(.002)	(.0017)	—	—	—	(.0021)	(.002)	(.002)
in months	.00047	.0008				.0002	.0005	.0005
GARM *1,000	.07	.6557						
size of the gov't	(.063)	(.52)	—	—	—	—	—	—
military in '000s	.22	.18						
GDP *1,000	−.0745	−.0733	−.0182	−.049	−.0435	−.00956	−.037	−.0241
(real income per	(.082)	(.069)	(.016)	(.06)	(.063)	(.089)	(.09)	(.082)
capita, PPP)	−.023	−.02	—	—	—	−.0029	−.01	−.0068

TABLE 8 *(cont.)*

Dependent Variable: WAREND2 *(Did the War End for Two Years?)*	*Model 1 Only Ethnic Wars Assume Exogeneity*	*Model 2 All Wars Assume Exogeneity*	*Model 3*[b] *2SLS Linear Prob. Assume Endogeneity*	*Model 4*[c] *2S Probit w/ Rivers & Vuong Test*	*Model 5*[d] *Bivariate Probit rho-Test for Exogeneity*	*Model 6 Probit w/ Partition Interaction Only Ethnic Wars*	*Model 7 Same as Model 6, but for the 5-Year Period*	*Model 8 Probit w/ Partition Interaction All Wars*
:H ethnic	-.048	-.045*				-.081***	-.066***	-.055**
heterogeneity	(.032)	(.025)	—	—	—	(.029)	(.022)	(.022)
index	-.015	-.0127				-.025	-.021	-.015
:HLPOP EH	.0027	.0024				.004**	.003***	.003**
interacted with	(.0019)	(.0015)	—	—	—	(.0018)	(.0013)	(.0014)
log of population	.00084	.0007				.0014	.0011	.0008
VARTYPE		-.265						
ethnic war or	—	(.34)	—	—	—	—	—	—
not?		-.072						
TEH						.013	-.004	.011
interaction:	—	—	—	—	—	(.012)	(.016)	(.010)
partv2*eh						.0039	-.001	.003
TLCOST						.311	.22	.244
interaction:	—	—	—	—	—	(.283)	(.311)	(.265)
partv2*logcost						.0961	.069	.0687
TARMY						-.0015	-.0025*	—
interaction:	—	—	—	—	—	(.001)	(.001)	
partv2*garm						-.0004	-.0008	
THPART	—	—	—					-3.82
interaction:	—	—	—	—	—	—	—	(3.46)
partv2*wartype								-.91

TABLE 8 *(cont.)*

Dependent Variable: WAREND2 *(Did the War End for Two Years?)*	*Model 1 Only Ethnic Wars Assume Exogeneity*	*Model 2 All Wars Assume Exogeneity*	*Model 3*[b] *2SLS Linear Prob. Assume Endogeneity*	*Model 4*[c] *2S Probit w/ Rivers & Vuong Test*	*Model 5*[d] *Bivariate Probit rho-Test for Exogeneity*	*Model 6 Probit w/ Partition Interaction Only Ethnic Wars*	*Model 7 Same as Model 6, but for the 5-Year Period*	*Model 8 Probit w/ Partition Interaction All Wars*
CONSTANT	1.51	2.25**	1.06***	1.69	1.62	1.75	2.36*	2.29***
	(1.07)	(.86)	(.305)	(1.03)	(1.03)	(1.167)	(1.26)	(.916)
Rho coefficient	—	—	—	—	.093	—	—	—
					(.443)			
Observations	78	122	78	78	78	78	73	123
Log-likelihood	−35.955	−57.399	—	−40.916	−63.633	−35.895	−33.913	−59.211
Pseudo-R^2	0.2399	0.1824	0.1619	0.1350	—	0.2412	0.2409	0.1602
Correctly Classified	79.50%	81.14%	—	—	—	76.92%	78.08%	77.23%
Reduction in Error	37.56%	26.32%	—	—	—	19.72%	28.50%	11.05%

*** significant at the .01 level; ** significant at the .05 level; * significant at the .10 level

[a] Reported are coefficients, standards errors (in parentheses), and marginals (dF/dx), in that order.

[b] Marginals reported only for models 1 and 2 since they are the ones used for inferences. Model 3 goodness-of-fit: F(4, 39) = 4.61. Instrumented: partition; Instruments: wardur garm gdp eh ehlpop.

[c] The first stage regression of model 4 has a Wald chi2(7)=36.38, a log-likelihood= -2.727, and a Pseudo-R^2 = 0.4606. All regressors except GARM are significant at the 5 percent level.

[d] Instruments as in models 3 and 4. All variables except GARM significant at the 5 percent level in the partition selection equation. The selection model's Wald chi2(11) = 50.02.

the basis of previous research on war termination. A regressor is used only if at least one other scholar has identified it as significant for war termination and peace building.[65]

The results of models 1 and 2 of Table 8 are quite robust to specification tests. A first interesting finding is that, though not significant in either model, partition is positively correlated with ethnic war recurrence in model 1 (it is negatively correlated with WAREND).[66] It is, however, positively correlated with WAREND if we look at the entire population of wars. In fact, among ideological nonethnic wars, partition is a significant determinant of war termination, which implies that nonethnic partitions are more stable and peaceful than ethnic partitions (although these results are driven by the very few cases of nonethnic partition).[67]

The human cost of the war in both models 1 and 2 is negatively correlated with war ending, confirming that the greater the number of people injured by the war, the harder it is to build peace.[68] War outcomes are the most significant and robust variable, though the OUTCOME2 variable is hard to interpret (Table 9 below explains these effects more clearly).[69]

[65] Following are the explanatory variables and the researchers who identified their importance: war duration, Mason and Fett (fn. 33); size of the government's military, Mason and Fett (fn. 33); war outcomes, Licklider (fn. 32) and Walter (fn. 39); ethnic heterogeneity and population size, Collier and Hoeffler (fn. 45); deaths and displacements, Licklider (fn. 32) and Doyle and Sambanis (fn. 45); income per capita, Collier (fn. 45); major power involvement, Singer and Small (fn. 32); foreign intervention, Regan (fn. 33); third-party and UN peace operations, Doyle and Sambanis (fn. 45); democracy, Collier, Elbadawi, and Sambanis (fn. 30) and Håvard Hegre et al., "Towards a Democratic Civil Peace? Opportunity, Grievance, and Civil War, 1816–1992" (Paper presented at the World Bank Conference on the Economics of Political and Criminal Violence, Washington, D.C., February 16–22, 1999). I also controlled for the number of land borders, the decade the war started, and the cold war. I could not include too many of these variables together, given my small data set and the collinearity of these variables.

[66] The Russia-Chechnya case causes the negative sign, and partition is positively correlated with WAREND if I drop that observation. The nonsignificance of partition, however, does not change by deleting that observation. I also sequentially deleted several other cases (e.g., Cyprus 1963/67, Somalia, Tajikistan, India), and the substantive results did not change. The results are also robust to using TYPELICK and TYPESTF instead of WARTYPE to identify ethnic wars.

[67] The opposite argument (without much empirical support) is made in Chaim Kaufmann, "Intervention in Ethnic and Ideological Civil Wars: Why One Can Be Done and the Other Can't," *Security Studies* 6 (Autumn 1997).

[68] The sign of LOGCOST in models 1 and 2 is negative, indicating that the higher the human cost, the greater the probability of war recurrence. This may seem counterintuitive: why would more costly wars lead to new wars? Would not great human cost discourage war recurrence? That reasoning is correct and it is reflected in my findings on war duration (see below), which verify the war-weariness hypothesis. However, controlling for this finding, the probability of war recurrence should be expected to be greater as the human costs of the war increase. These costs measure war-generated hostility and create grievances that may manifest themselves in future conflict. Further, the greater the human and economic cost of the war, the lower should be a country's human capital and the lower the state's capacity to resume normal operations.

[69] OUTCOME2 is a categorical variable denoting whether the war ended in a truce, rebel victory, government victory, or peace settlement. It is highly significant in both models 1 and 2, though interpreting its sign is not straightforward. By disaggregating OUTCOME2, I found some of its components highly correlated with partition. Thus, entering them independently in the regression would increase collinearities. These correlations also generate sufficient concern over the possible endogeneity of partition when both OUTCOME2 and partition are included in the model.

TABLE 9
FIRST DIFFERENCES OF THE PROBABILITY OF WAR RECURRENCE FOR ETHNIC AND ALL CIVIL WARS[a]

Only Ethnic Wars *Difference of*	*Estimate of Mean Change in Prob(warend2)=1*	*Standard Deviation of Estimate*	*95% Confidence Interval*	
PARTITION from 0 to 1	-.027	.174	-.374219	.3025837
EH from 25th to 75th percentile	-.601	.322	-.9731405	.2372383
EHLPOP & EH from 25th to 75th percentile	.002	.095	-.1844085	.1865729
OUTCOME2 from truce to treaty	.119	.027	.064833	.1706
LOGCOST from 25th to 75th percentile	-.182	.097	-.3737424	.0045121

All Civil Wars *First Difference of*	*Estimate of Mean Change in Prob(warend2)=1*	*Standard Deviation of Estimate*	*95% Confidence Interval*	
PARTITION from 0 to 1	.088	.129	-.1755012	.3330738
EH from 25th to 75th percentile	-.621	.269	-.9458801	.1065577
EHLPOP & EH from 25th to 75th percentile	-.082	.067	-.2188984	.0452573
OUTCOME2 from truce to treaty	.098	.025	.0507963	.1499487
LOGCOST from 25th to 75th percentile	-.246	.088	-.416339	-.0723445
WARTYPE from ethnic to ideological	-.094	.123	-.3369971	.1517012

[a] Changes in X-variables are noted in the first column. All other variables are held constant at their means. Some of these probability changes are not statistically significant, as can also be seen in Table 8.

Generally, the more conciliatory the outcome of the war, the greater the probability that the peace will last. War duration is positively correlated with no war recurrence, but it is only significant in model 2. This result is consistent with the findings of Mason and Fett and weakly supports

the "war weariness" hypothesis.[70] The size of the government's military is also positive but nonsignificant (it is significant only at the 10 percent level in model 2).[71] Real GDP per capita is negatively correlated with war ending, though this association is not significant.[72]

The impact of ethnic heterogeneity on war recurrence is quite interesting. Ethnic heterogeneity (EH) and its interaction term with the log of population (EHLPOP) are not significant in models 1 and 2, despite the emphasis on ethnicity in the civil war literature. Moreover, EHLPOP has a positive sign, indicating that if the ethnic groups are large enough to protect themselves against domination, the risk of war recurrence is smaller. This result is consistent with Fearon and Laitin's analysis of the potential for interethnic cooperation and is robust to different measures of ethnic division (the ELF index) and different model specifications.[73] As I added several interaction terms between partition and other variables in models 6–8, I found a significant negative association between ethnicity and lasting peace but also a significant positive association between peace and larger ethnic groups. Other authors have found that ethnolinguistic and religious fractionalization is a significant determinant of the initiation of civil wars, but my findings suggest that civil war initiation and war recurrence are different phenomena with respect to the role of ethnicity and that ethnic diversity is not as detrimental to peace as many tend to assume.[74]

Models 1 and 2 present short-term results. To see if these results also hold in the medium term, I reestimated models 1 and 2 both five and ten years after the end of the war: the model's fit was roughly the same. However, while no major differences occur in the five-year model, in the ten-year model, war duration is no longer significant and partition is highly significant and negatively correlated with war termination.[75] In models 6–8 I added several interaction terms between partition and LOGCOST, WARTYPE, EH, and GARM, but no appreciably important associations emerged. In model 8 the coefficient of ETHPART—denoting

[70] Mason and Fett (fn. 33). I thank Russ Leng for pointing this out to me.

[71] By contrast, Mason and Fett (fn. 33) find a significant negative correlation.

[72] This finding may be due to selection effects, as I mentioned earlier, but it may also confirm the revenue-seeking economic model of civil war in Collier (fn. 45); and Collier and Hoeffler (fn. 45); the model looks at GDP per capita as a proxy for "lootable" resources, which would increase the risk of war.

[73] James Fearon and David Laitin, "Explaining Interethnic Cooperation," *American Political Science Review* 90 (December 1996).

[74] See, e.g., Collier, Elbadawi, and Sambanis (fn. 30).

[75] This could be due to missing observations. The ten-year model was estimated on fifty cases. The coefficient and robust standard error of partition were −1.137 and .627, respectively, yielding a z-statistic of −1.813 ($P>|z| = 0.070$) and a model log-likelihood = −23.460, with a Wald chi2(8) = 32.14 and a Pseudo $R^2 = 0.3199$.

ethnic partition—*is negatively but not significantly correlated with war termination.*

These findings build a strong case against partition theory. Before accepting them, I should address a methodological point with potentially substantive implications. My inferences may depend on the assumption that partition is exogenous to war recurrence, an assumption that makes sense, since war recurrence is observed after the occurrence of partition. It may be, however, that these two variables are jointly determined by a common set of explanatory variables. Thus, there is concern over the possible endogeneity of partition. I therefore estimated several models assuming partition is endogenous and I also tested its exogeneity (see the appendix for technical details and for an explanation of the statistical concept of endogeneity). I present the results of three models that address the issue of endogeneity in Table 8.

I estimate a parsimonious linear probability model treating partition as an endogenous variable (model 3). The model uses several variables that are not correlated with war recurrence (based on the previous estimations) as instruments for partition and reduces the number of exogenous variables in the structural equation to reduce noise in the system of equations. The model has good fit and is well specified. Despite this estimation adjustment, however, partition is not at all significant and continues to have a negative sign. The other key variables (LOGCOST, OUTCOME2, GDP) behave in much the same way as they did in the previous models.

In models 4 and 5 I formally tested (and could not reject) the null hypothesis that partition is exogenous. Model 4 is a two-stage probit model, estimated following Madalla and using the Rivers and Vuong exogeneity test (which was originally designed for continuous endogenous explanatory variables).[76] The exogeneity test consists of a t-test of the estimated coefficient of the predicted residual (RES) of the first-

[76] G. S. Maddala, *Limited Dependent and Qualitative Variables in Econometrics* (Cambridge: Cambridge University Press, 1983); D. Rivers and Q. Vuong, "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models," *Journal of Econometrics* 39 (1988). This method is almost identical to Kenneth A. Bollen, D. K. Guilkey, and T. A. Mroz, "Binary Outcomes and Endogenous Explanatory Variables: Tests and Solutions with an Application to the Demand for Contraceptive Use in Tunisia," *Demography* 32 (February 1995). The two-stage probit model produces inefficient standard errors, though the efficiency loss is small. Rivers and Vuong derive the formula that gives the correct variance-covariance matrix, but their procedure is designed for continuous endogenous right-hand-side variables. A methodological discussion and Monte Carlo simulation results reporting the properties of this estimator in small samples are found in Michael Alvarez and Jennifer Glascow, "Two-Stage Estimation of Non-Recursive Choice Models," *Political Analysis* (forthcoming). For a political science application of this method, see Michael Alvarez and L. Butterfield, "The Resurgence of Nativism in California? The Case of Proposition 187 and Illegal Immigration," *Social Science Quarterly* (forthcoming). A discussion of the two binary variable case can be found in Madalla (p. 246).

stage reduced form equation of partition. The estimated coefficient of RES is .122 and its standard error is .956, so we cannot reject the null hypothesis that partition is exogenous. To confirm this result, I estimated model 5—a bivariate probit model with sample selection (which is a full-information maximum likelihood model)—using the same instruments for partition and the same exogenous variables as in models 3 and 4.[77] The correlation coefficient *rho* (see Table 8) is .093 and its standard error is .443. A Wald test of the hypothesis that *rho* = *0* yields chi2(1) = .043, which does not allow us to reject the exogeneity of partitions at the 0.8346 level. Thus, more efficient estimates are obtained from models 1 and 2 and we can rely on the previously discussed inferences.

Table 9 provides some easily interpretable results of the estimated change in the probability that there will be no war recurrence as a result of changes in key explanatory variables (these changes are also known as first differences). I reestimated models 1 and 2 from Table 8 and simulated (with one thousand repetitions) their parameter estimates using the CLARIFY software.[78] Using those estimates, I obtained the first differences reported.[79] I report the mean and standard deviation of the estimated change in probability (the standard deviation reflects the significance levels of the explanatory variables as in Table 8). Note that the probability of no war recurrence (peace) is slightly reduced by a partition after an ethnic war, although it is increased by 8.8 percent when we use the entire population of civil wars (though this result is not statistically significant). Among the statistically significant effects, it is worth reporting that if the war ends in a treaty instead of a truce, the probability of no war recurrence increases by 11.9 percent in ethnic wars and 9.8 percent in all wars. Finally, the more costly the war, the less stable the peace. By varying LOGCOST from its 25th to its 75th percentile, the probability of no war recurrence drops by 18.2 percent in ethnic wars and 24.6 percent in all wars.

To conclude this section, the evidence does not support the assertion that partition significantly reduces the risk of war recurrence. Hence

[77] For the case of two binary variables, this can be estimated as a seemingly unrelated bivariate probit model with a selection effect. The exogeneity test in this model consists of a Wald test of *rho*, the estimated coefficient of the correlation of the error terms in the structural and reduced-form ("first-stage") equations.

[78] Michael Tomz, Jason Wittenberg, and Gary King, "Clarify: Software for Interpreting and Presenting Statistical Results," version 1.2.1 (June 1, 1999). See also idem, "Making the Most of Statistical Analyses: Improving Interpretations and Presentation" (Paper presented at the annual meeting of the American Political Science Association, Boston, 1999).

[79] Note that these estimates may differ slightly in replications since I did not fix the number seed used to randomly select samples for the simulations.

Does Partition End Ethnic Violence Short of War?

The third critical hypothesis in support of partition is that the physi separation of ethnic groups will reduce residual, low-level ethnic vi lence (that is, violence short of war). In this section I test this hypot esis empirically and find that only under carefully specified conditio does partition reduce lower-level violence. In most situations partitio will have a negligible effect on residual violence.

The dependent variable in the models estimated in this section is NOVIOL, which takes the suffixes 2 and 5 when it is measured two and five years after the end of the war, respectively. NOVIOL is binary and it is coded 1 if there is no residual violence after the end of the war and 0 otherwise. It is coded based on available information and other data sets that code armed conflict short of war and/or politicides and genocides (a table with details on the coding of each case is included in the online data set). I use two versions of the dependent variable, NOVIOL2 and NOVIOL5, for two and five years after the end of the war.

I start by regressing NOVIOL2 on the explanatory variables from the war-recurrence model, which does not produce many significant results and has low explanatory power. Partition is positively but nonsignificantly associated with an end to lower-level violence (model 1 in Table 10). Given the model's low classification success, it appears that war recurrence and residual violence are substantially different phenomena, so I changed the model's specification in regressions 2–6. First, I disaggregated the OUTCOME2 variable and focused on the signing of treaties, which in model 2 increased the significance level of partitions by doubling its coefficient and reducing its standard error. Treaty is also significant and positive, suggesting that treaties generally do signal the parties' intentions to reconcile their differences. We also find that EHLPOP is significant and positive.[80] Both results seem important. The

[80] This result complements Bates's (fn. 47) findings on the relationship between ethnicity and political violence in Africa. Bates studies the relationship between ethnicity and economic modernization—urbanization, education, and the rise of per capita income, as well as political participation. Focusing on forty-six African countries from 1970 to 1995, he tests the relationship between ethnicity and political violence at several levels (not just wars). He measures the size distribution of ethnic groups, linguistic diversity, and the presence of an ethnic minority at risk, based on the work of a number of other researchers, and shows that the relationship between ethnicity and violence is complicated. Bates finds that "controlling for the impact of other variables, [linguistic diversity] associates with higher levels of violence . . . the size of the largest ethnic group enters quadratically; when the coefficient for the linear term is significant, so too is the coefficient for the quadratic. But as the size of the largest ethnic group grows, the level of violence initially decreases, but then increases; by contrast, the level of protest

s to support partition theory, though the latter seems to
t. These results are essentially the same for the five-year pe-
OL5; see model 3 in Table 10).[81]
ls 3 and 4 are differently specified to facilitate sensitivity analy-
he previous results and to achieve higher classification success.
3—discussed previously—applies to the five-year period and has
than double the classification success of the two-year model. It
s that partition is not significant for ending residual violence. In
del 4 I added the variable INTENSE, which measures the war's inten-
y and which is positive and extremely significant. Very intense wars
em either to provide no further appetite for violence or to eradicate
ll resistance early on, so no residual violence is necessary. The GARM
ariable is not significant in any of the regressions and neither is economic development (proxied by electricity consumption per capita).[82] In model 4 we find that residual violence is significantly reduced by long and bloody wars, but again partition is not significant. Further, the partition variable is extremely fragile to small specification changes. Contrast models 4 and 5, which differ only in that model 4 controls for LOGCOST and model 5 for LOGDEAD (that is, not for displaced persons). The coefficient of the partition variable drops from .67 to .27 (the marginal effect drops from .26 to .10). The other variables are much more stable.

Model 6 is the best-performing model specification; it controls for civil wars in the same country during the previous ten years. That variable is extremely significant and suggests that residual violence is much more likely if the country has had a history of civil war recurrence. This effect is mitigated, however, in the case of long and intense wars, which discourage residual violence. Model 6 has substantially higher classification success than the previous models, but, in general, the models of

initially increases, but then falls" (p. 25). In Bates's analysis, it is extreme polarization that is most associated with violence. My finding that greater heterogeneity and more sizable ethnic groups reduce violence in postwar states is therefore compatible with Bates's result and complements it. I should note however, that Bates's results may not be generally applicable to non-African countries, given that African countries have a generally higher mean level of ethnic heterogeneity and this implies a selection effect *if the results are to be applied widely to non-African countries*. In my data set, for example, the mean and standard deviation of the ethnic heterogeneity index for African countries is 68.82 and 34.58, respectively, whereas for non-African countries it is 50.5 and 31.73. The ethnic heterogeneity index used was created by Tatu Vanhanen, "Domestic Ethnic Conflict and Ethnic Nepotism: A Comparative Analysis," *Journal of Peace Research* 36, no. 1 (1999).

[81] Ten years after the end of the war partition has a negative correlation with an end to low-grade violence, but this result may be an artifact of missing (right-censored) observations (only fifty-five observations are available for the ten-year period).

[82] Thus, development levels and military strength are more relevant with respect to war recurrence than with respect to low-level violence.

TABLE 10

PROBIT REGRESSIONS OF NO-VIOLENCE

(PART ASSUMED EXOGENOUS)[a]

Dependent Variable: NOVIOL2 *(No Violence for 2 Years?)*	*Model 1* *Ethnic War*	*Model 2* *Ethnic War*	*Model 3* *Ethnic War*	*Model 4* *Ethnic War*	*Model 5* *Ethnic War*	*Model 6* *Ethnic War*
Constant	.783	1.37	1.14	1.55	.586	1.95**
	(.868)	(.924)	(.868)	(.987)	(.978)	(.93)
PART	.298	.668*	.632	.672	.272	.685
did the war result	(.402)	(.387)	(.411)	(.418)	(.371)	(.441)
in partition?	.116	.26	.248	.262	.106	.268
LOGCOST natural	–.144*	–.167*	–.125	–.195**		–.208**
log of deaths &	(.08)	(.087)	(.085)	(.094)	—	(.088)
displacements	–.055	–.064	–.048	–.075		–.080
INTENSE				16.14***	12.05***	21.47**
human cost per	—	—	—	(4.83)	(3.715)	(9.04)
capita per month				6.26	4.69	8.34
LOGDEAD					–.104	
natural log of	—	—	—	—	(.097)	—
deaths					–.04	
OUTCOME2	.276**					
outcome of the war	(.123)	—	—	—	—	—
—four outcomes	.106					
TREATY		.883*	.992**	.869	.756	.955*
was there a treaty	—	(.534)	(.482)	(.559)	(.526)	(.557)
signed?		.341	.380	.335	.294	.367
WARDUR	.0021	.0026	.0028	.004**	.0027	.0047**
duration of war	(.0017)	(.0017)	(.0018)	(.002)	(.0019)	(.0022)
in months	.0008	.0010	.0011	[illegible]		

TABLE 10 *(cont.)*[a]

Dependent Variable: NOVIOL2 *(No Violence for 2 Years?)*	*Model 1 Ethnic War*	*Model 2 Ethnic War*	*Model 3 Ethnic War*	*Model 4 Ethnic War*	*Model 5 Ethnic War*	*Model 6 Ethnic War*
GARM*1,000 size	–.317	–.302	–.173	–.29	–.28	–.174
of the government	(.27)	(.28)	(.22)	(.308)	(.26)	(.31)
military	–.1	–.1	–.067	–.11	–.11	–.067
ELECTRIC*1,000	.055	.009	–.040	–.053	–.0195	–.134
electricity consumption p.c.; kilowatt hours	(.117)	(.124)	(.13)	(.13)	(.15)	(.12)
	.02	.0036	–.015	–.02	–.0076	–.052
EH	–.017	–.02*	–.027*	–.032**	–.035**	–.024*
ethnic heterogeneity	(.013)	(.011)	(.015)	(.013)	(.014)	(.012)
Index	–.006	–.0077	–.011	–.012	–.013	–.0092
EHLPOP	.00097	.001*	.0013*	.0017***	.0019***	.0014**
ethnic heterogeneity	(.00069)	(.0005)	(.00074)	(.00068)	(.00073)	(.00062)
* log of population	.00037	.00042	.0005	.00068	.00074	.00055
PW10						–.85***
civil war in	—	—	—	—	—	(.34)
previous ten years?						–.323
Observations	77	77	72	77	77	77
Log-likelihood	–45.952	–45.608	–43.020	–43.713	–45.983	–40.606
Pseudo-R^2	0.1209	0.1274	0.1258	0.1637	0.1203	0.2231
Classification success	63.63%	61.03%	65.27%	67.53%	66.23%	71.42%
Reduction in error	20%	14.25%	46.79%	28.55%	25.69%	37.11%

*** significant at the .01 level; ** significant at the .05 level; * significant at the .10 level

[a] Coefficients, standards errors reported (in parentheses), and marginals (dF/dx).

residual violence do not fit the data as well as the models of war recurrence discussed earlier. Thus, the lack of significance of the partition variable may be a function of omitted variables that could be identified in further theoretical research. At the same time model 6 confirms the results discussed earlier with respect to the positive impact of ethnic diversity, if ethnic groups are large (EHLPOP variable).

In sum, I can point to only very weak evidence in support of the hypothesis that partitions help end low-level ethnic violence (in model 2 and some variants). More importantly, the positive impact of partitions seems fragile and extremely dependent on whether or not the war ended in a treaty, on the war's intensity, on the number of people displaced by the war, and on the number and size of ethnic groups.

To confirm that these results are not influenced by the potential endogeneity of partitions to low-grade violence, I tested once again for exogeneity. I estimated model 4 using a 2SLS linear probability model using many different combinations of instrumental variables to see whether the coefficient of partition would become significant if endogeneity is assumed. I used as instruments only variables that were not significantly associated with low-level violence in the previous regressions. I could not, however, find any results that would increase my confidence in the significance of the partition variable.[83]

To summarize, I find only weak support for the hypothesis that partitions are significant for an end to low-grade ethnic violence after civil war. Models supporting that hypothesis are very sensitive to small specification changes. Models controlling for the potential endogeneity of partition are not robust, have poor fit to the data, and are quite sensitive to the choice of instrumental variables.

CONCLUSION

Population movements to partition states during or after civil war are coerced, painful, and costly, and they may sow the seeds of future conflict. It is therefore imperative that international policy toward partition

[83] Results are available from the author. Finding good instruments in this data set has proven notoriously difficult. I did not change the specification of model 4; rather I just added and dropped instruments and considered other variables as potentially endogenous (specifically, WARDUR, LOGCOST, TREATY, ENERCAP). The instruments I used in various combinations were GEO, BORDER, EH, GDP, URBST (urban population at the start of the conflict), GARM, INTERVEN (was there an external intervention?), MAJOR (was there a major power involved?). I found only one permutation that made partition significant, but when I estimated this model using a bivariate probit estimator, I found that the error terms of the structural and reduced-form first equation were perfectly correlated (*rho* = 1), which indicates either that there was too much noise in the system or that the distribution is not a bivariate normal.

be informed by rigorous, empirically verified arguments, rather than by untested theory. In this paper, I have provided a host of empirical tests, starting with an empirical inquiry into the determinants of war-related partition. I have found that partitions are more likely after costly ethnic/religious wars, after rebel victory or truce, and in countries with better-than-average socioeconomic conditions. Partitions are more likely where ethnic groups are large; they are less likely to occur as the degree of ethnic heterogeneity increases.

My analysis has also shown that the differences between ethnic and nonethnic wars with respect to war termination and partition are small. The relationship between the degree of ethnic heterogeneity and the need for partition is not as straightforward as partition theorists assume.[84] The finding that partition does not significantly prevent war recurrence suggests, at the very least, that separating ethnic groups does not resolve the problem of violent ethnic antagonism.

These findings lead me to formulate a new hypothesis for future research: if border redefinition is in the cards after civil war (or before war), then the strategy of supporting ethnic diffusion by combining rather than partitioning large ethnic groups may be worth pursuing. Thinking along these lines could be very useful for Africa in particular, given the persistent concern about the "unnaturalness" of Africa's borders and the recent debate about redefining those borders in the hope of reducing the incidence of civil wars.[85] Partition theorists would argue for the partition of warring African states into a multitude of ministates each of which is composed of a single ethnicity group. Even if this solution reduces the incidence of internal war, it will almost certainly increase the incidence of international war. Based on the empirical findings of this paper, I would put forth a rival hypothesis: if borders can be credibly and securely redrawn, then combining several large ethnic groups in a larger multiethnic state may reduce the probability of new wars. It would be useful for further research to try to clarify any threshold effects that may be associated with the size of ethnic groups as they affect the likelihood of violent secessionist activity.

No doubt this proposal is difficult, perhaps infeasible. Skeptics, however, should at least acknowledge that the political dangers of ethnic diversity are imperfectly understood and that, despite the practical

[84] To cite Horowitz (fn. 2), 135: "Is there any reason to believe that the more pronounced the cultural differences that exist between groups, the greater the ethnic conflict? There has been no shortage of offhanded assertions that cultural differences engender ethnic conflict. But . . . systematic statements of the relationship are more difficult to find."

[85] See, e.g., Jeffrey Herbst, "Responding to State Failure in Africa," *International Security* 21 (Winter 1996–97).

difficulty of my proposal, it is at least based on a more nuanced understanding of the relationship between ethnicity and political violence. Indeed, much of the best recent research on political violence consistently points to the fact that ethnic diversity need not generate violence. Economic studies of the occurrence of civil war in a sample of 161 countries have found a parabolic relationship between ethno-linguistic and religious fractionalization and the onset and duration of civil wars.[86] Work in political science, too, has steadily reinforced this point.[87] Bates, for example, casts "doubt upon a . . . tenet of conventional wisdom: that ethnic diversity promotes violence." He finds that the "red zone" for violent conflict occurs "when an ethnic bloc may be sufficient in size to permanently exclude others from the exercise of power."[88] Thus, enhancing ethnic diversity while strengthening political institutions can be beneficial. So why partition states to reduce ethnic diversity?

Perhaps there are other benefits to be derived from partition—democratization, for example. The jury is still out on this question. My empirical analysis shows that partitions may have such an effect, but this effect may be due to the prewar institutions of civil war–torn countries; such institutions may be at least as important as partition in determining the democratic future. And the process of democratization itself may harbor dangers and cause violence.[89] Thus, international policy aimed at preventing war recurrence should promote institution building and socioeconomic development before war occurs in the first place, rather than supporting partition after war occurs.

Partition, as we have seen, does not help reduce the risk of war recurrence. Partitions are in fact positively (though not significantly) associated with recurrence of ethnic war. The probability of a new war rises in tandem with the human toll of the previous war and with nondecisive outcomes to the war. War recurrence is also positively, though not significantly, associated with GDP per capita[90] and with ethnic heterogeneity, though as ethnic groups become larger, new wars tend to

[86] Collier, Elbadawi, and Sambanis (fn. 30).

[87] Laitin and Fearon (fn. 73).

[88] My findings need not agree entirely with Bates's (fn. 47), since our samples and research questions differ. I have not measured the size of the largest ethnic group, which is critical in his argument, but I do find that greater ethnic heterogeneity within the context of a larger population reduces the risk of war recurrence and residual violence. This is consistent with Bates.

[89] Bates (fn. 47), 28. The dangers of the process of democratization, as opposed to the end goal of democracy, should not be underestimated. On the potential of regime transitions, including democratic transitions, to create civil war, see Hegre et al. (fn. 65).

[90] This correlation between GDP and war recurrence may seem counterintuitive since we saw that higher GDP is correlated with higher democracy, which is negatively correlated with civil war. The impact of GDP is ambiguous because GDP can be a proxy both for the country's overall development level, which should be positively associated with peace, and for "loot," inciting new wars; see Collier and Hoeffler (fn. 45).

become less likely.[91] Negotiated settlements, a strong government army, and a lengthy previous war all reduce the probability of war recurrence. Thus, if the international community's interest lies in preventing new civil wars, it could manipulate some of these significant variables toward desirable goals. It could, for example, take steps to enhance the government's military and support decisive war outcomes.[92] Or it could support the negotiation of peace treaties, which reduce the threat of new violence. Again, what the international community should not do to prevent future wars is promote partition.

One benefit that can come from partition is the reduction of residual low-level ethnic violence. Certainly, it follows that if ethnic groups divide into ethnically homogeneous territorial units, the risk of ethnic conflict declines. I have found that this result depends critically on the number of displaced people, as well as on the way the previous war ended and on the country's longer war history. The probability of low-level violence is shaped by some of the same variables that influence war recurrence, but it also is a function of different determinants. The previous discussion of how to prevent war therefore applies to residual violence, but with some modifications. To reduce residual violence, it is important to prevent war recurrence, as patterns of large-scale violence over time seem to encourage lower-level violence.

Fine-tuning a war-to-peace transition is a difficult task because peace-building strategies can often backfire. For example, promoting economic growth may assist democratization and promote peace, but it can also lead to new wars by expanding the potential economic gains from a new rebellion. Strategies to support the government's prewar institutions and its military may also achieve peace, but they may do so at the expense of justice.

Muddling through this difficult terrain, I would propose an empirically derived strategy for resolving ethnic wars. This strategy demands action by the international community, which must promote democracy as its number one conflict-prevention strategy. If violence does erupt, its priority should be to facilitate a negotiated settlement, as well as to integrate and downsize the government's military. According to my empirical analysis, these strategies can be effective. If border redefinition is a viable option—and it should be an option only if it does not

[91] Despite the fact that this study looks only at cases where war has already taken place, the results on the impact of ethnic heterogeneity are compatible with studies of initial war occurrence, in that ethnic heterogeneity seems to increase the risk of war at first, but as the degree of heterogeneity increases, that risk declines. In my analysis, however, this effect has not been statistically significant.

[92] Such a strategy, however, might indirectly support political repression, so it must be carefully and selectively [illegible]

assist one party at the expense of another—then ethnic integration rather than ethnic partition may be a winning strategy. In addition to having the potential for greater success than partition, this strategy is also not loaded with subjective and arbitrary assumptions about the necessity for ethnically pure states and about the futility of interethnic cooperation. Only in the most extreme cases may partition be necessary, indeed inevitable. Those cases must be handpicked on the basis of political analysis of regional and global constraints, the history of the preceding war, and the special traits of the society in question. More research on this topic will help pinpoint the benefits and the dangers of partition under different conditions. What this study has suggested is that, on average, partition may be an impossible solution to ethnic civil war.

Appendix: Testing for the Possible Endogeneity of Partitions

A number of explanatory variables are correlated with both partition and no war recurrence. This raises concerns about possible endogeneity of partition in models of war recurrence. This appendix explains the problem of endogeneity to which I refer in the text and explains how the problem can be addressed.

Endogeneity results if a system of simultaneous equations has correlated error terms. Here, we have a system of two equations of partition (P) and war recurrence (W):

$$P_i = \alpha + \beta_1 M_i + \beta_2 X_i + \varepsilon_i \qquad (1)$$
$$W_i = \gamma + \beta_3 P_i + \beta_4 M_i + \beta_5 Z_i + \mu_i \qquad (2)$$

where equation 1 determines the incidence of partition (P) and equation 2 determines war recurrence (W). M_i is a vector of common variables in the two equations and β_1 is the vector of coefficients of M_i; X_i is a vector of instrumental variables (correlated with partition but not with war recurrence) and β_2 is a vector of their coefficients; Z_i is a vector of variables that are correlated with war recurrence and β_5 is a vector of their coefficients. Endogeneity stems from the possible correlation between ε_i and μ_i (both of them disturbance terms with mean zero and no correlation to the other explanatory variables). The presence of common explanatory variables in equations 1 and 2 (that is, the variables in M_i) implies that ε_i and μ_i may be correlated, in which case, a simple probit regression would produce biased parameter estimates.

To resolve this problem, we must first test for endogeneity, and if we detect it, we must estimate a model that corrects for it as in the two-stage least squares model for linear regression. This problem is harder when the dependent variable is binary. The two-state probit used by Bollen, Guilkey, and Mroz, following Rivers and Vuong, is a model that has been shown to generate consistent parameter estimates with small efficiency loss. This model has attractive properties.[93] The procedure is similar to 2SLS: first, estimate the reduced form of equation 1 using a probit regression, obtaining predicted values (PHAT) of the dependent variable (PART).[94] Then, compute the error of that prediction (RES) and plug the PHAT into equation 2, replacing PART. PHAT is uncorrelated with the disturbance term in equation 1. Then, estimate equation 2 using a probit regression with robust standard errors. The resulting coefficient estimates are consistent asymptotically inefficient.[95] However, evidence from small-sample Monte Carlo simulations has shown that the efficiency loss is very small.[96]

To determine whether this process should be used, a simple exogeneity test can be applied. This involves a t-test of the estimated coefficient of the prediction error (RES), added as a regressor along with the actual value of PART in equation 2 in the model of war recurrence.[97] If RES is nonsignificant, we cannot reject the null hypothesis of exogeneity of partitions and we should rely on the simple probit estimates. Applying this method to the no-war-recurrence model, I find that more efficient results are reached by models that assume exogeneity. This is confirmed by estimating a bivariate probit model of war recurrence that provides efficient standard errors (estimated by maximum likelihood) and a different test of exogeneity.

[93] The choice of estimator depends on the number of observations, the degree of identification of the model, the number of potentially endogenous variables, and the goodness of fit of the first-stage equation. See Bollen, Guilkey, and Mroz (fn. 76).

[94] Use of this method is strictly based on attaining an R^2 no smaller than 10 percent in equation 1. In my case, the R^2 was higher than 30 percent. Further, Bollen, Guilkey and Mroz (fn. 76) discuss evidence from Monte Carlo simulations that suggest that the models' identification must be less than 75 percent for the two-step probit estimator to be preferable to the simple probit (i.e., the overlapping variables in the two equations must be fewer than three-fourths of the total number of right-hand-side variables in the structural equation).

[95] T. Amemiya, "The Maximum Likelihood and the Nonlinear Three-Stage Least Squares Estimator in the General Nonlinear Simultaneous Equation Model," *Econometrica* 45 (1978).

[96] See Bollen, Guilkey, and Mroz (fn. 76); and G. Tauchen, "Diagnostic Testing and Evaluation of Maximum Likelihood Models," *Journal of Econometrics* 30 (1985). Formulas to compute efficient standard errors can be found in Maddala (fn. 76), and a method to estimate the asymptotically efficient covariance matrix when the model is overidentified has been developed by Amemiya (fn. 95). Alvarez and Butterfield (fn. 76).

DEGREES OF DEMOCRACY
Some Comparative Lessons from India

By PATRICK HELLER*

INTRODUCTION

ONE of the most remarkable developments of the late twentieth century has been the number of countries that have made the transition from authoritarian to democratic rule. Even if, following Linz and Stepan,[1] one employs an exacting definition of democratic consolidation, there are far more countries today in which democracy is the only game in town than was the case just fifteen years ago. This development has in turn produced an empirically and theoretically rich literature on democratic transitions and consolidation. But if we have learned much about the conditions and processes under which the transition to democracy takes place, we have only just begun to scratch the surface of the equally challenging question of the *effectiveness* of democracy in the posttransition period. As Linz and Stepan note, beyond the consolidation of democracies there is room to

> improve their quality by raising the minimum economic plateau upon which all citizens stand and by deepening political and social participation in the life of the country. Within the category of consolidated democracies there is a continuum from low to high quality democracy; an urgent political and intellectual task is to think about how to improve the quality of most consolidated democracies.[2]

This paper takes up this challenge by drawing on the case of India, which is of particular significance for theories of democratic deepening on two counts. First, the general picture of Indian democracy stands as a reminder that there is no linear progression to democracy. Much as

* This paper was first presented in the Politics and Identity Workshop at Columbia University. I would like to thank my colleagues at Columbia University—Anthony Marx, Doug Chalmers, Harrison White, and David Stark—as well as Peter Evans, Judith Tendler, T. M. Thomas Isaac, Atul Kohli, Ronald Herring, Peter Lange, Glenn Adler, and Robert Keohane for their feedback. I am especially grateful to Charles Tilly, Dietrich Rueschemeyer, Charles Kurzman, Deborah Barrett, and the anonymous reviewers for their extensive comments and insights. The research for this paper was supported by the American Institute of Indian Studies.

[1] Juan Linz and Alfred Stepan, *Problems of Democratic Consolidation* (Baltimore: Johns Hopkins University Press, 1996).

[2] Ibid., 6.

the robustness of India's democratic institutions has been rightfully celebrated, the effectiveness of those institutions is increasingly in doubt. Fifty-three years of almost uninterrupted democratic rule has done little to reduce the political, social, and economic marginalization of India's popular classes. Second, a more disaggregated picture reveals that within the unitary institutional domain marked by the boundaries of the Indian nation-state, there are degrees of democracy or, as Guillermo O'Donnell has put it, differences in the intensity of citizenship.[3] India's posttransition history has produced multiple trajectories of democratization. Taken together these observations inform the central theoretical argument of this paper: in order to understand the conditions under which democracy can be deepened, we need to develop accounts of democratization that explore the dynamic interactions between institutions and social processes.

I define posttransition democratic deepening as a process under which the formal, effective, and substantive dimensions of democracy become mutually reinforcing. In much of the late-developing world social and economic conditions have conspired to limit the capacity of subordinate groups to effectively exercise their rights and to secure substantive gains. This has often produced a vicious cycle in which the ineffectiveness of formal democracy produces increased social tensions, which in turn trigger autocratic political responses and "movements of rage."[4] In India the increasing incidence of caste and communal violence, the criminalization of politics, the spread of corruption, and the rise of ethnic-chauvinist and communitarian parties all point to a crisis of the democratic state.[5] Within India there are, however, important exceptions.[6] Atul Kohli, for instance, has argued that the continuous

[3] O'Donnell, "On the State, Democratization and Some Conceptual Problems: A Latin American View with Glances at Some Postcommunist Countries," *World Development* 21, no. 8 (1993).

[4] Huber, Rueschemeyer, and Stephens provide a discussion of why a "vicious cycle" of interaction between these three dimensions of democracy—which they label formal, participatory, and social—is symptomatic of new democracies. Evelyne Huber, Dietrich Rueschemeyer, and John D. Stephens, "The Paradoxes of Contemporary Democracy: Formal, Participatory, and Social Dimensions," in Lisa Anderson, ed., *Transitions to Democracy* (New York: Columbia University Press, 1999).

[5] Mehta, for example, notes that "there is a widespread sense that the state and its laws can be manipulated to serve particular interests, that official decisions are guided by no principle but expediency, and that even this uncertain restraint on the arbitrary use of power is wearing thin." Pratap Mehta, "India: Fragmentation amid Consensus," *Journal of Democracy* 13, no. 1 (1997). See also Sugata Bose and Ayesha Jalal, eds., *Nationalism, Democracy and Development* (New Delhi: Oxford University Press, 1999), 59.

[6] Although this is hardly surprising in a federal system marked by such a high degree of socioeconomic heterogeneity, most treatments of Indian democracy focus on the national picture. When democracy is examined at the subnational level, the emphasis has almost always been on variations in the political party system, not on democratic deepening. State-level studies moreover remain more concerned with [illegible]

rule of a disciplined, left-of-center party in West Bengal—the Communist Party of India, Marxist (CPM)—has produced both order and some success in alleviating poverty.[7] The case has also been made that because of different histories of caste reform movements, democracy in South India is deeper than in the North, marked by more political participation and significant achievements on the social development front.[8] The most distinct departure from the general pattern, however, is the case of the southwestern state of Kerala (population thirty-one million). Here, as I hope to show, the procedural, effective, and substantive elements of democracy have evolved in a virtuous cycle, resulting in the political and economic integration of subordinate classes. Often cited as an exceptionally successful case of social development, substantive democracy in Kerala—which has included successful land reforms, poverty reduction, and social protection measures—has been tied to a long history of social mobilization and effective public intervention.[9] In contrast to the transition literature and an older literature on political modernization, both of which emphasize the difficulties developing countries face in managing high levels of popular mobilization, the case of Kerala suggests that under certain conditions, organized societal demands and democratic governance can be mutually reinforcing. And if the case of Kerala helps us understand the conditions under which democratization flourishes, it also allows us to reconstruct why the same conditions have not obtained for the rest of the country. This reconstruction in turn takes us beyond the analyses of elite decision making and institutional design that have informed much of the transition literature.[10] It brings us to causal accounts that in the

[7] Kohli, *The State and Poverty in India* (Cambridge: Cambridge University Press, 1987); idem, *Democracy and Discontent: India's Growing Crisis of Governability* (Cambridge: Cambridge University Press, 1990).

[8] Ashutosh Varshney, "Is India Becoming More Democratic," *Journal of Asian Studies* 59 (February 2000).

[9] On all the key measures of quality of life, Kerala has achieved levels that rival those of the West and are decades ahead of national averages. For a classic early study, see United Nations, *Poverty, Unemployment and Development Policy: A Case Study of Selected Issues with Reference to Kerala* (New York: Department of Economic and Social Affairs, 1975). For the most recent comparison with India, see Jean Drèze and Amartya Sen, *India: Economic Development and Social Opportunity* (Delhi: Oxford University Press, 1995). Important studies that highlight the role of social movements include Richard Franke and Barbara Chasin, *Kerala: Radical Reform as Development in an Indian State* (San Francisco: Food First Development Report no. 6, 1989); K. P. Kannan, *Of Rural Proletarian Struggles: Mobilization and Organization of Rural Workers in South-West India* (Delhi: Oxford University Press, 1988); and V. K. Ramachandran, "On Kerala's Development Achievements," in Jean Drèze and Amartya Sen, eds., *Indian Development* (Delhi: Oxford University Press, 1996).

[10] Most notably, see Guillermo O'Donnell, Philippe Schmitter, and Laurence Whitehead, eds., *Transitions from Authoritarian Rule: Prospects for Democracy*, 4 vols. (Baltimore: Johns Hopkins University Press, 1986). In a review of the transition literature, Shin writes that the establishment [illegible]

tradition of Barrington Moore's classic work on the social origins of democracy highlight the ways in which historically conditioned dynamic relations among broadly constituted social actors drive not only the *making* of democracy but also the *deepening* of democracy.[11] Such an analysis, in particular, helps us excavate the critical role played by subordinate classes.[12]

THE DEMOCRACY DEBATE

The debate on democratic transitions has understandably focused on the installation of electoral, constitutional, and procedural institutions, with the unit of analysis invariably the nation-state. While useful for typologizing regimes and differentiating democracy from authoritarianism, focusing on formal national-level institutions provides only limited analytical leverage for conceptualizing democratic deepening. Because institutions and politics are relational and configurational, their attributes are never perfectly isomorphic either horizontally across different policy arenas or vertically from one level of the state to another. As the state radiates out from its geographic and functional core, its authority and its effectiveness fluctuate dramatically. Much as state-society theorists have recently called for disaggregating the state,[13] we need to disaggregate democracy. As conventionally defined, formal democracy is marked by universal suffrage, regular and competitive elections,

conscious choices among various types of democratic constitutions, and electoral and party systems." Doh Chull Shin, "On the Third Wave of Democratization: A Synthesis and Evaluation of Recent Theory and Research," *World Politics* 47 (October 1994), 139.

[11] Moore, *Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World* (Boston: Beacon Press, 1966).

[12] If the first generation of analysts of democratic transitions highlighted the role of elite political actors (see fn. 10), a second generation has looked beyond the moment of the negotiation of transitions to more historical explanations in which the role of subordinate actors assumes greater explanatory weight. See, for example, Ruth B. Collier and David Collier, *Shaping the Political Arena: Critical Junctures, the Labor Movement, and Regime Dynamics* (Princeton: Princeton University Press, 1997); Ruth B. Collier and James Mahoney, "Adding Collective Actors to Collective Outcomes: Labor and Recent Democratization in South America and Southern Europe," *Comparative Politics* 24, no. 3 (1997); Anthony Pereira, *The End of the Peasantry: The Rural Labor Movement in Northeast Brazil, 1961–1988* (Pittsburgh, Pa.: University of Pittsburgh Press, 1997); Gay W. Seidman, *Manufacturing Militance: Workers' Movements in Brazil and South Africa, 1970–1985* (Berkeley: University of California Press, 1994); Deborah Yashar, *Demanding Democracy: Reform and Reaction in Costa Rica and Guatemala, 1870s–1950s* (Stanford, Calif.: Stanford University Press, 1997); and Glenn Adler and Eddie Webster, "Challenging Transition Theory: The Labor Movement, Radical Reform, and Transition to Democracy in South Africa," *Politics and Society* 23 (March 1995). For the most important treatment of how subordinate actors, and in particular working classes, have contributed to democratic deepening, see Dietrich Rueschemeyer, Evelyne Huber Stephens, and John D. Stephens, *Capitalist Development and Democracy* (Chicago: University of Chicago Press. 1992)

[13] Joel Mi[illegible]

accountability of state apparatuses to elected representatives, and legally codified and enforced rights of association.[14] An *effective* democracy is one in which democratic practices have spread throughout society, governing not only relations between states and citizens but also public relations between citizens. Functionally and geographically the degree of public legality in many formal democracies remains severely constrained. In such democracies, notes O'Donnell, "the component of democratic legality and, hence, of publicness and citizenship, fades away at the frontiers of various regions and class, gender and ethnic relations."[15] Public spaces disappear to be replaced by areas of privatized power. Local institutions and officials are colonized by bosses, chiefs, dons, or caciques. Patrimonialism, clientelism, and coercion eat away at democratic authority. Thus we must look beyond the macroinstitutional level of parliaments, constitutions, and elections. And we must investigate instead the intermediate- and local-level institutions and consultative arenas located in the interstices of state and society where "everyday" forms of democracy either flourish or founder. We need, in other words, a political sociology of democracy, one that specifically recognizes that a working democracy must be an effective democracy.

An effective democracy has two interrelated characteristics—a robust civil society and a capable state. A free and lively civil society makes the state and its agents more accountable by guaranteeing that consultation takes place not just through electoral representation (periodic mandates) but also through constant feedback and negotiation. Civil society is critical to democratic performance because it extends the scope and style of claim making beyond the formal interest representation that defines political society. Social movements, associations, and unions raise new issues and mobilize new actors. In doing so they not only provide a counterbalance to more bureaucratic and aggregated forms of interest representation, but they also create new solidarities, which in many instances specifically challenge existing inequalities and hence help democratize society itself. The key point here is that the health of a democracy is measured as much in the qualitative nature of its social patterns of association as in the formal character of its institutions. And while these two variables condition each other—associational patterns are conditioned by institutional environments, and institutional responsiveness is conditioned by associational vitality—the interaction can be positively reinforcing just as it can be mutually corrosive.

[14] Huber, Rueschemeyer, and Stephens (fn. 4), 168.
[15] O'Donnell (fn. 3), 1361.

The capacity of the state is also central to the effectiveness of democracy. Procedural guarantees of civic and political rights, including rights of association and free speech, do not automatically translate into the effective exercise of democratic rights. Citizenship is not a right, it is a relation. Where inequalities between social categories are so pronounced as to create extraconstitutional forms of binding authority (clientelism, patriarchy, caste subordination), the exercise of citizenship is subverted. As theorists of civil society have long argued, its associational qualities emerge only when it is doubly differentiated from the state and from primary social groupings (families, kinship groups, lineages). A precondition for the effective exercise of civic and political rights requires a state capable of securing the even, uniform, and rational-legal enforcement of public authority. Individuals and groups must be protected from arbitrary state action but also from forms of social authority that might constrain or impinge upon their civic and political liberties. And creating public spaces that are protected from nondemocratic forms of authority requires far more than writing constitutions and holding officials accountable. It marks a fundamental shift in the distribution and locus of what Weber called "legitimate domination" from society to state. Given the contested and unfinished process of state formation in much of the developing world, the writ of legally enforced public authority remains limited, producing a low-intensity form of citizenship.[16] This problem is tied to both the infrastructural and the authoritative limits of state power. Infrastructurally, the apparatuses of the state—the police, the judiciary, the educational system—are simply cast too thinly and too unevenly to enforce and provide for citizen's rights. Authoritatively, the state's legitimate realm of domination (constitutionally prescribed arenas in which its authority is binding and backed by coercion) is contested and weakened by countervailing sources of authority.

Finally, in the context of the developing world the distinction between formal and substantive democracy has to be taken more seriously. Because the Western trajectory of democratization passed, as T. H. Marshall argued,[17] through incremental stages of the civil, the political, and the social, most political scientists have treated the formal and substantive aspects of democracy as discrete and sequential phenomena. Yet while the distinction between process and outcome is heuristically useful, in the context of the developing world there are compelling reasons not to treat these attributes of democracy in analytical isolation.

[16] Ibid.

First, a strong case can be made that the persistence of acute social inequities compromises the basic logic of associational autonomy that informs the classical liberal claim for defending procedural democracy on its own merits.[18] As Weyland notes, "Abject poverty forces many people into clientelist bonds with elites who offer minimal benefits and protection in exchange for obedience and political support—that is for an abdication of citizen rights."[19] In much of the developing world—and especially in the so-called informal sector—economic relations are to a great extent reproduced through social and political forms of domination. The intimate nexus between these domains, which exists at both the structural and the individual level, reduces the effective utility of formal political rights to subordinate groups. As Vilas notes succinctly, "Institutional rules are less important in these situations than personal arrangements and connections based on reciprocities, real or symbolic, explicit or implicit."[20]

Second, in the postcolonial and postsocialist world, where the franchise was extended wholesale rather than incrementally, democratic legitimations were infused from the outset with substantive demands for distributive justice. Because peasants and workers were politically empowered at a much earlier stage of economic development than were their European counterparts, the tension between citizenship rights and property rights has always been especially acute. But in most democracies of the developing world the predominant pattern of subordinate-class incorporation has been clientelism or populism, rather than economic integration through the expansion of social citizenship, a pattern that has invariably exacerbated social tensions and threatened democratic stability.[21] In Latin America unresolved conflicts between propertied elites and the redistributive demands of organized labor or peasant movements have repeatedly produced authoritarian responses. In the East European context Przeworski et al. have argued that democratic consolidation is critically tied to the effectiveness with which elected governments can manage the social costs of market transitions. Linking the question of the legitimacy and viability of democratic rule to substantive outcomes, Przeworski et al. argue that "regardless of their

[18] For an important discussion of this point, see Jonathan Fox, "The Difficult Transition from Clientelism to Citizenship: Lessons from Mexico," *World Politics* 46 (January 1994).

[19] Kurt Weyland, *Democracy without Equity: Failures of Reform in Brazil* (Pittsburgh, Pa.: University of Pittsburgh Press, 1996), 5.

[20] Carlos Vilas, "Democracy and the Whereabouts of Participation," in Doug Chalmers, Scott Martin, and Kerianne Piester, eds., *The New Politics of Inequality in Latin America* (New York: Oxford University Press, 1997), 19.

[21] Nicos Mouzelis, "Modernity, Late Development and Civil Society," in John A. Hall, ed., *Civil Society: Theory, History and Comparison* (Cambridge, UK: [illegible]

age, democracies persist whenever all the major political forces find that they can improve their situation if they channel their demands and their conflicts within the democratic institutions."[22]

The substantive outcomes against which a democratic regime is judged are of course historically contingent. Comisso, for example, argues in her review of the track record of democratization in Eastern Europe that substantive democracy has assumed three different forms—liberal (sustaining a free enterprise economy), national (preserving a particular national characteristic), and egalitarian (promoting equity).[23] Most irreducibly, however, effective democracies give rise to redistributive pressures, as indeed classical theorists from Locke to Marx were quick to recognize. Przeworski makes this same point with characteristic clarity: "If the median voter is decisive, and if the market-generated distribution of income is skewed toward lower incomes (as it always is), then majority rule will call for an equality of incomes."[24] Where such interests fail to be aggregated or where the state fails to be responsive, democratic consolidation itself is threatened. As Linz and Stepan comment: "If a democracy never produced policies that generated government-mandated public goods in the areas of education, health, transportation, some safety net for its citizens hurt by major market swings, and some alleviation of gross inequality, democracy would not be sustainable."[25]

In sum, any understanding of democratic deepening must recognize the complicated interplay of the formal, effective, and substantive dimensions of democratic rule. Formal civic and political rights mean ittle if they are not backed by the authoritative power of the state or ıre routinely suspended or transgressed by extrademocratic powers. And effective rights of association and representation will necessarily give rise to substantive tests of the legitimacy of democratic rule.

THE LIMITS OF EFFECTIVE DEMOCRACY IN INDIA

n any federal parliamentary system, and especially one that has been rected on an extremely heterogenous social landscape, the quality of

[22] Adam Przeworski et al., *Sustainable Democracy* (Cambridge: Cambridge University Press, 1995),). In an earlier work Przeworski framed the question of transitions in static political equilibria terms, ncluding that successful transitions are built on elite pacts that are "inevitably conservative, economically and socially." Quoted in Adler and Webster (fn. 12), 84.

[23] Ellen Comisso, "Is the Glass Half Full or Half Empty? Reflections on Five Years of Competitive litics in Eastern Europe," *Communist and Post-Communist Studies* 30, no. 1 (1997)

[24] Adam Przeworski "Th...

democracy is bound to vary dramatically. Within the boundaries of the Indian nation-state the formal institutional parameters of democratic rule can be held constant. Variations instead have to be located along the effective and substantive dimensions of democracy. These are dynamic relationships that have to be explored through a configurational analysis. Kerala lies at one end of the range of possibilities and represents one of many possible paths to democratic deepening in India. In order to tease out its configurational specificity and to identify the causal historical dynamics at work, the following section attempts to provide both context and comparative leverage by exploring the general pattern of posttransition democratization in India.

India's democratic institutions have withstood the test of time and the test of a fissiparous society. The basic procedural infrastructure of democracy—specifically the constitution and guarantees of the rights of association, the separation of powers, and regular and open elections at both the national and the state level—has become firmly entrenched.[26] At a minimum, despite infamous recent episodes of communal and caste violence, democratic institutions have not only helped forge a nation from multiple nationalities but have also institutionalized acceptance of the uncertainty of rule that comes with competitive elections.[27] The authoritarian episode of 1975–77 notwithstanding, the prospects of a democratic reversal in India are remote. India's dominant class factions, proprietary and professional, support democracy, if for no other reasons than that they have benefited so handsomely from the largesse of India's democratic politics of patronage.[28]

The effectiveness of India's democratic institutions is an altogether different matter. Throughout vast regions of India the exercise of citizenship rights, even in the limited political sense of the term, is circumscribed by the persistence of traditional forms of social control. With more than half of India's rural households depending on landlords for access to land or labor, clientelistic ties remain key to the sur-

[26] There are a total of twenty-five states in India, fifteen of which have populations surpassing fifteen million (hereafter the "major" states). Indian states have their own legislatures and executives and under India's federal constitution enjoy a wide range of powers and responsibilities, including independent sources of revenue collection (primarily sales taxes) and a wide range of development functions.

[27] One could simply point to border areas—the Northeast region, Kashmir, and Punjab—where separatist struggles have led to the suspension of basic democratic rights as the extreme cases of undemocratic practice. These, however, are in effect areas where the legitimacy of the nation-state itself is contested. It is variation within the boundaries of the consolidated and legitimated Indian nation-state that are of concern here.

[28] Pranab Bardhan, *The Political Economy of Development in India* (New York: Basil Blackwell, 1984).

vival strategies of subordinate groups.[29] With its ritualized exclusions and deeply ingrained hierarchical relations, the caste system has inscribed these material inequalities with a degree of social and cultural control that has few parallels. Low levels of literacy and discriminatory treatment by upper-caste-controlled state institutions have further limited the associational autonomy of lower castes and classes. The corollary of this picture is the predominance of fragmented sovereignty. The reach and authority of the juridical and democratic state ends—or more accurately is transfigured—where the writ and power of local strongmen and their caste-based followings begin. In a pattern that closely resembles both the Brazilian and the African cases, extrademocratic sources of authority not only resist but also colonize and privatize state power.[30] Local notables routinely dominate local institutions, including village governments, schools, cooperative societies,and the development bureaucracy. The permeability of state authority is most dramatically exposed by the existence of private caste armies (especially in Bihar) and elite control over local police forces. On these counts, the general picture of Indian democracy bears a striking resemblance to Putnam's description of the uncivic regions of southern Italy:

> Public life in these regions is organized hierarchically, rather than horizontally. The very concept of "citizen" here is stunted. From the point of view of the individual inhabitant, public affairs is the business of somebody else—*i notabili*—"the bosses," "the politicians" —but not me. . . . Political participation is triggered by personal dependency or private greed, not by collective purpose.[31]

This is not to say that democratic institutions in India have been altogether lifeless. The past two decades have witnessed an erosion of traditional clientelist politics. Formal and competitive democracy in India has undermined the legitimacy of traditional social authority, spawned a whole new generation of political entrepreneurs, and created spaces in which new groups have been successfully mobilized. But the political forces that have emerged are more rooted than ever in social cleavages.[32]

[29] In 1982, 66.6 percent of rural households fell below the marginal ownership category of 2.5 acres. Together they accounted for 12.2 percent of total land ownership, less than the 14.3 percent share of the top 1 percent of households. C. T. Kurien, *The Economy: An Interpretive Introduction* (New Delhi: Sage Publications, 1992), 321.

[30] For Brazil, see Weyland (fn. 19). For Africa, see Mahmood Mamdani, *Citizen and Subject: Contemporary Africa and the Legacy of Late Colonialism* (Princeton: Princeton University Press, 1996).

[31] Robert Putnam, *Making Democracy Work: Civic Traditions in Modern Italy* (Princeton: Princeton University Press, 1993), 115.

[32] Pradeep Chhibber, *Democracy without Associations: Transformation of the Party System and Social Cleavages in India* (Ann Arbor: University of Michigan ...

The basis for mobilization has shifted from patronage to identity populism. The pattern has varied greatly from state to state, but the trend has been one of a marked increase in the political saliency of essentialized identities of caste, religious community, and ethnicity (subnationalism). That this pattern has affected every state in India underscores the significance of two broad developments. The proximate cause is the decline of the Congress Party and its catchall electoral politics. The deeper structural cause lies in the well-documented social and redistributive failures of the Indian state.[33] The electoral dominance of the Congress Party was sustained by vast and vertically organized patronage networks that held together a wide range of interests and groups. The resulting pattern of state-society engagement that Herring has called "embedded particularism"[34] nourished rent-seeking interests at the expense of the state's capacity to provide public goods and institutional reform. These developmental failures have unleashed new sources of social conflict.

As elections have become more competitive and more groups have been brought into the political arena on their own terms, patronage has become increasingly tied to identity politics. The demand for government quotas and special privileges, whether of majority or minority communities, now dominates claim making. This explosion of narrow demands has triggered a frantic zero-sum scramble for preferential treatment that Bardhan has aptly described as "equal-opportunity plundering by all interest groups."[35] Subjected to this chaotic chorus of particularistic claims, governing coalitions have become increasingly opportunistic and unstable. Thus scholars ranging from a neo-Huntingtonian to a neo-Marxist persuasion have noted a severe erosion of the state's autonomy and the eclipse of its developmental mandate.[36]

In this political climate of populism and organizational fragmentation, encompassing political formations have been the exception to the rule. Labor unions have rarely extended beyond the protective confines of the organized sector (large factories and public employees) and in many instances have become little more than vehicles for the political ambitions of local bosses. Farmers' associations have been dominated

[33] See Drèze and Sen (fn. 9, 1995 and 1996); and Kohli (fn. 7, 1987).

[34] Ronald Herring, "Embedded Particularism: India's Failed Developmental State," in Meredith Woo-Cumings, ed., *The Developmental State* (Ithaca, N.Y.: Cornell University Press, 1999).

[35] Pranab Bardhan, "Sharing the Spoils: Group Equity, Development, and Democracy" (Manuscript, Department of Economics, University of California, Berkeley, 1997), 16.

[36] See Bardhan (fn. 28); Lloyd I. Rudolph and Susanne Hoeber Rudolph, *In Pursuit of Lakshmi: The Political Economy of the Indian State* (Chicago: University of Chicago Press, 1987) [illegible]
24).

by the interests of large farmers. With the reach of public legality circumscribed by the power of local elites, lower-class and lower-caste efforts to organize around economic issues (outside of Kerala and West Bengal) have invariably been defeated.[37] And while the past decade has witnessed an explosion in NGOs, in part as a response to failures of the developmental state, their coverage remains spotty. And outside of some well publicized cases, their capacity to scale-up and impact public policy has been limited.

In the absence of cohesive lower-class organizations, mass politics has little programmatic content. Varshney, for example, has convincingly shown that in the national election of 1996 economic reforms were a nonissue. Instead, mass political discourse was dominated by "expressions of India's identity politics [which] have led to mass mobilization, insurgencies, riots, assassinations, desecrations and destructions of holy places. In popular perceptions, the significance of identities has been far greater than the implications of economic reforms."[38] The relationship between the primacy of identity politics and the unorchestrated character of subordinate politics is captured by the rise to power of the Hindu-nationalist BJP (Bharatiya Janata Party). Against the general involution of the political party system, the relative cohesiveness of the BJP marks a post-Congress political reconsolidation of upper-caste, middle-, and lower-middle-class domination, with significant support from business interests. The BJP's electoral appeal is rooted in a nationalist politics of order that explicitly denies deep societal cleavages of class, caste, and ethnicity through the construction of an "imagined community." As Bose notes, the ideological potency of the BJP lies in "the majoritarian myth of Hinduvata . . . with its millenarian vision of an India which has resolved all its problems, political conflicts and social contradictions through an affirmation of the organic unity of a common 'Hindu' identity."[39] The claim of inclusiveness is belied by the explicit exclusion of Muslims and the implicit defense of the caste-stratified status quo. As for subordinate classes, they remain splintered between Muslim, dalit (untouchable), and backward caste political formations, in which the centrality of demands for reservation

[37] Paul R. Brass, *The Politics of India since Independence* (Cambridge: Cambridge University Press, 1994), 334.

[38] Ashutosh Varshney, "Mass Politics or Elite Politics? India's Economic Reforms in Comparative Perspective" (Paper presented at the Conference on India's Economic Reforms, Center for International Affairs and Harvard Institute of International Development, Harvard University, Cambridge, December 13–14, 1996), 3.

[39] Sumantra Bose, "Hindu Nati[illegible]

of government and educational positions serves the narrow interests of the upwardly mobile elites of these communities. Against this backdrop of political fragmentation, the capacity of democratic institutions to aggregate interests and in particular to address pressing distributional dilemmas is more in doubt than ever. The politics of social citizenship, as Mehta has remarked, are conspicuous by their absence.[40] Thus, much as in the case of Brazil, political fragmentation has frustrated the equity-enhancing potential (and promise) of democracy.[41]

This failure of Indian democracy to give effective voice to substantive demands has locked in a vicious cycle that is eroding the very legitimacy of democratic governance. With the state's failure to provide basic services and a modicum of protection from market swings, ordinary citizens are seeking security in traditional networks and have become increasingly susceptible to the politics of scapegoating. Much as O'Donnell has noted for Latin America, in a climate of insecurity and uncertainty, politics becomes a desolidarizing affair.[42] The increasing prevalence of privatized strategies and the politics of "sauve qui peut" is reflected in the widely documented "criminalization"[43] of politics, the proliferation of large-scale corruption scandals that have tarnished most parties, and the upsurge of sectarian and castist violence. To complete the vicious circle, a state that has failed to secure stable and workable ties to society has come to rely increasingly on administrative power and coercion rather than on democratic participation.[44] Describing this erosion of public legality, one of the most seasoned scholars of Indian politics writes of "an increasingly pervasive Hobbesian state of disorder, unpredictability and fear of violence among ordinary people in the rural areas of India."[45]

Democracy in Kerala

Kerala shares the same formal democratic institutions that are found in the rest of the nation. Most of its fiscal resources and much of its economic health depend on the central government. Its administrative

[40] Mehta (fn. 5), 64.

[41] Weyland (fn. 19).

[42] O'Donnell (fn. 3), 1365–67.

[43] To cite one example: "According to the Chief Election Commissioner, 180 of Uttar Pradesh's [India's most populous state] Members of the Legislative Assembly (MLAs) have criminal cases pending against them, at least 52 of which involve 'heinous crimes' such as rape and murder." Jean Drèze and Haris Gazdar, "Uttar Pradesh: The Burden of Inertia," in Drèze and Sen (fn. 9, 1996), 108.

[44] See Rajni Kothari, *State against Democracy: In Search of Humane Governance* (Delhi: Ajanta Books, 1988).

structures are those of the center, and many of its highest ranking bureaucrats are recruited from a national pool. Yet on all key qualitative measures of democracy, Kerala clearly stands apart from the center and from all other Indian states. The effectiveness of its democratic institutions is reflected in the degree and scope of public legality. In Weberian terms, rational-legal authority has displaced traditional authority in the regulation of public life. Not only have individuals achieved greater autonomy and capacity in exercising their democratic rights, but subordinate groups have been successfully integrated into public politics. Moreover, democracy in Kerala has produced significant and measurable substantive outcomes, most notably important redistributive reforms and the expansion of the welfare state. These substantive outcomes are the result of the inclusionary and encompassing character of democratic life, but they have also helped reinforce the effectiveness of state authority and levels of political participation.

If democracy in Kerala works better than in the rest of India, it is in large part because individuals have been equipped with the basic human capabilities required of citizenship. Literacy in Kerala has reached 91 percent, compared with 49 percent for India as a whole. Not only have successive governments maintained the highest rates of educational expenditure in India, but in contrast to the national pattern they have pursued a strategy of mass education by prioritizing primary universal schooling over secondary education.[46] As a direct result, traditionally marginalized groups, most notably women and dalits, have acquired the basic social skills necessary for informed participation. This is concretely reflected in the highest per capita circulation of newspapers in India[47] and rates of electoral participation that run 15 to 20 percent higher than the national average. Associational life has been further strengthened by the provision of basic public goods. The provision of basic health care and subsidized food staples and the regulation of the labor market have reduced material dependencies and eroded traditional clientelistic networks. Increased associational autonomy, most notably for subordinate groups, is in turn reflected in the sheer density of civic organizations and the vigor of associational life. Keralites of all walks of life, it would seem, have an irresistible inclination to combine, associate, and organize.[48] There are a large number of

[46] See Ramachandran (fn. 9).

[47] Robin Jeffrey, *Politics, Women and Well-Being: How Kerala Became a "Model"* (Houndmills, Basingstoke, Hampshire: Macmillan Press, [illegible]

NGOs operating in the state, including the mass-based KSSP (Kerala Sastra Sahithya Parishad), which has achieved world renown for its efforts to "bring science to the people." Kerala's caste self-help and social uplift societies have a long history of active civil engagement. Overlapping an extensive public school network is a network of private and semiprivate schools sponsored by communal and caste organizations. The result is a school in every village and nearly universal primary school enrollment.[49] The same pattern is found in the health sector.

It is not sufficient to positively correlate associational life with democratic deepening, however. Just as a vibrant civil society can promote trust and cooperation, it can also promote particularism that fosters rent-seeking lobbies and exclusionary identities. In a context of pervasive social inequality and highly skewed distribution of political resources and influence, the openness of Indian democracy has produced highly differentiated associative capacities and has given full play to oligarchical coalitions or, in the Rudolphs' phrase, demand-group politics.[50] The narrowness of interest aggregation has in turn subverted the public good, and the Indian state has become an instrument for creating and reproducing a "network of advantage distribution."[51] And while India's caste system and religious communities, evolving as they have in a pluralist political system, have promoted popular involvement in public life and spawned a dense configuration of self-help, cultural, educational, youth, and women's organizations, they have done so on the strength of parochial (and often exclusionary) identities and organizational structures that retain important elements of patriarchal authority and hierarchy. This pattern of associational life has in turn shaped the logic of political society. The dominant political formations in most Indian states are increasingly either communal parties (the BJP), regionalist parties (the AIADMK in Tamil Nadu and the Telugu Desam in Andhra Pradesh), or loosely organized coalitions of caste-based factions (the Janata Dal).

India (Ithaca, N.Y.: Cornell University Press, 1999); and idem, "Social Capital as Product of Class Mobilization and State Intervention: Industrial Workers in Kerala, India," *World Development* 24, no. 6 (1996). In a detailed study that focuses on the North Kerala city of Calicut, Varshney concludes that "[m]uch like Tocqueville's America, Calicut is a place of 'joiners.' Associations of all kinds—business, labor, professional, social, theater, film, sports, art, reading—abound." Ashutosh Varshney, *Civic Life and Ethnic Conflict: Hindus and Muslims in India* (New Haven: Yale University Press, forthcoming), manuscript pp. 130–31.

[49] The retention rate in primary schools—the percentage of children having entered primary school who complete the fifth grade—is 82 percent in Kerala as against 26 percent for India as a whole. Myron Weiner, *The Child and the State in India* (Princeton: Princeton University Press, 1991), 174.

[50] Rudolph and Rudolph (fn. 36).

Caste and community in Kerala continue to be a powerful basis of social identity and civic engagement. But in the realm of politics and in the expression of public authority, these forms of association have been subordinated to broader aggregations, in particular, class-based organizations. By all accounts Kerala has the highest levels of unionization in the country, and unlike the national pattern, the presence of unions is not limited to the formal sector of the economy. In addition, large numbers of workers in the informal sector—including in the beedi, construction, coir, and cashew industries—are organized.[52] Most notably, the largest union in the state is the CPM-affiliated KSKTU, the one-million-member union of predominantly lower-caste agricultural workers. Kerala's mass organizations of women, students, and youth—sponsored by all the political parties—also play an active role in the state's political life. The CPM's mass organizations alone claim a membership of over 4.7 million. The state's network of cooperative societies, which are controlled by political parties not communities, is the most extensive in the country. And in the political arena the basic cleavage has been along class lines, opposing a coalition of right-wing parties organized around the Congress against a coalition of left-wing parties organized around the Communist Party of India, Marxist.[53] These two coalitions have more or less alternated in power, and consistently thin margins of victory point to a relatively stable distribution of political support.

The effectiveness of Kerala's democratic institutions is best measured by the extent to which they have successfully managed social and economic tensions. To a much greater degree than in any other Indian state, Kerala's political history has been shaped by open and organized class conflict. Because Kerala is one of India's poorer states and until recently its most densely populated, distributional conflicts have been especially acute. Recurrent waves of organized agitation have moreover been orchestrated by a Communist Party that for much of its history was wedded to a strategic line of extraparliamentary mass struggle. During the highly conflictual 1960s and 1970s governments were short-lived, rural protest was endemic, and rates of industrial unrest (as measured in strikes) were the highest in India. Kerala's social structure is moreover marked by significant cleavages. Its caste system was historically among the most rigidly stratified in India, and it has the

[52] For an extended discussion, see Heller (fn. 48, 1999), chap. 6.

[53] The Communist Party of India (CPI) was unified until 19[illegible]

largest minority concentrations of Christians and Muslims of any Indian state, with each group constituting roughly 20 percent of the population. The major caste and communal groups have powerful and active associations, and Christian and Muslim political parties have played key roles in Kerala's coalition politics. In Huntingtonian models of order such high levels of interest organization and political conflict are associated with problems of governance and/or civic disorder.[54] At no time, however, have Kerala's social and economic cleavages and its extraparliamentary politics threatened democratic governance. Class conflict in Kerala has not produced the armed Naxalite and other revolutionary groups that dot the Indian countryside, and Kerala's militant unions, while confrontational, have not become embroiled in the kind of organized violence and criminal networks that increasingly characterize organized labor in the rest of the country. Moreover, having institutionalized lower-class interests, the CPM abandoned the politics of class struggle in the 1980s in favor of a social democratic strategy of class compromise, and labor militancy fell dramatically.[55] Similarly, though Kerala's castes and communities are well organized, instances of sectarian violence and caste violence have been rare.[56] The question of reservations (quotas) for "other backward castes" that has repeatedly triggered violence and commanded the political spotlight since the Mandel Commission report of 1990 has had virtually no impact in Kerala. Most notably, the resurgence of Hindu majoritarianism in Indian national politics has had a negligible impact in Kerala.[57]

That civic harmony has been maintained in a climate of highly organized and politicized social forces and against a backdrop of low economic development points to the efficacy of Kerala's democratic institutions in mediating and absorbing conflict. At the macrolevel, a highly competitive electoral arena, the alternation in power of two bipolar political fronts, and the mobilizational capacity of opposition groups have all contributed to heightening the accountability of the state. At the meso and microlevels, a dense network of democratic in-

[54] Samuel Huntington, *Political Order in Changing Societies* (New Haven: Yale University Press, 1968).

[55] See Heller (fn. 48, 1999), chaps. 5, 6, and 7.

[56] Varshney (fn. 48) found that out of seventeen states, Kerala ranked the fourth lowest in the number of per capita deaths in urban communal riots (manuscript p. 112); this is based on a large-N data set on Hindu-Muslim violence—1950–95—jointly produced with Steve Wilkinson. Of the three states ranking below Kerala, the Punjab has been the site of sustained rural violence and neither Haryana nor Tamil Nadu has a sizable Muslim population.

[57] In Kerala's 1995 local government elections, the BJP captured only 3 out of 1,200 panchayats, and only 1 out of 26 municipalities. Kerala remains the only major [illegible]

stitutions and authorities is readily identifiable. In contrast to the general Indian picture in which district and village-level institutions are deeply enmeshed in local power configurations and often in the hands of landed elites or dominant castes, in Kerala a wide range of institutions including district councils, panchayats (local governments), student councils, and cooperative societies are hotly contested by the major political formations. Representative institutions have also directly penetrated economic life. Thus, one of the unique features of Kerala's economic scene is the role that voluntary tripartite bodies (industrial relations committees) representing labor, capital, and the state play in actively shaping and coordinating labor relations, industrial policy, and welfare programs across a wide range of industries, including agriculture.

Finally, there is a clear correspondence between high levels of political participation and government performance. Across virtually every public policy arena the effectiveness of state intervention in Kerala far surpasses the performance of any other Indian state and can be tied to demand-side pressures.[58] The provision of education, health care, and subsidized food has been characterized by universal coverage and comparatively corruption-free delivery and is a direct response to broad-based support (across all major political parties) for the extension of social rights. Redistributive measures, in particular, land reforms, labor-market regulation, and the extension of social protection schemes to informal sectors, have dramatically reduced levels of poverty, and can all be tied to specific episodes of sustained mobilization.[59] Most recently, the literacy campaign of 1991—the most successful in India—as well as a decentralization campaign initiated in 1996 (more below) were both the result of pressure exerted by left-leaning NGOs.

If democracy works better today in Kerala, it is because its citizens are active and organized and because horizontal forms of association prevail over vertical (clientelistic) forms of association. The resulting patterns of political participation have in turn favored encompassing demands that promote the public interest over narrow and fragmented demands for state patronage (rents). If civil society in most of India remains deeply embedded in social cleavages, why has civil society in Kerala been so decisively differentiated from predemocratic social structures?

Explanations for the growth of civil society generally fall into two camps. The most conventional is to see civil society as the expression of

[58] See Drèze and Sen (fn. 9, 1995).

[59] See fn. 9. For an extended comparative [illegible]

Ronald H [illegible]

autonomous private interests arising from the growth of a market economy. Alternatively, the quality of civic engagement has been interpreted as a cultural phenomenon of mutual respect and trust, its roots located in civic practices dating back, as in Putnam's influential argument about northern Italy, to the "mist of the dark ages."[60] Neither explanation fits the case of Kerala. Nineteenth-century Kerala was characterized by what is generally considered to have been the most rigid and severe caste system in the subcontinent[61] and an agrarian economy that, while exposed to early commercialization, was deeply rooted in labor-repressive institutions. A social system marked by an ascribed division of labor, ritualized and elaborate codes of degradation (untouchables were considered to be "unseeable" and lower-caste women forbidden to cover their breasts), and acute material dependencies was anything but fertile soil for civic republicanism. And despite significant commercialization of agriculture in the first half of the twentieth century, land in northern Kerala (Malabar) remained the monopoly of a parasitic class of Brahmin landlords, and capitalist agriculture in the South (Cochin and Travancore) relied heavily on bonded labor. Civil society in Kerala arose neither from deep civic traditions nor from the associational and gentlemanly impulses that Montesquieu attributed to commercial life. Instead, the birth of a vibrant and effective democracy in Kerala must be located in its political history of conflict and social mobilization, the interplay of these dynamics with the process of state building, and the resulting transformation of social structure.

THE SOCIAL ORIGINS OF DEMOCRACY IN INDIA AND KERALA

The divergent paths that democracy in Kerala and India has taken can be grasped only through a historical and configurational analysis. In untangling the processes at work I draw specifically on two related but discrete analytical clusters developed in Rueschemeyer, Stephens, and Stephens's comparative study of democratization.[62] Following in the tradition of Barrington Moore, Rueschemeyer, Stephens, and Stephens begin by highlighting the causal significance of class configurations. Specifically, they confirm the relationship between landed power and authoritarian rule and acknowledge the critical role of the bourgeoisie in promoting democratic reforms. They depart from Moore, however,

[60] Putnam (fn. 31), 180.

[61] Robin Jeffrey, *The Decline of Nair Dominance: Society and Politics in Travancore, 1947–1908* (New York: Holmes and Meier, 1976).

[62] Rueschemeyer, Stephens, and Stephens (fn. 12).

in emphasizing that the role of the bourgeoisie is a necessary but not sufficient condition of democratic consolidation. While the bourgeoisie has generally "been supportive of the installation of constitutional and representative government," it has been opposed to "extending political inclusion to the lower classes."[63] Constitutional rights and representative government served to protect private economic activity from the exactions and arbitrary interferences of absolutist states and monarchs. But private property represents both the basis and the limits of the bourgeoisie's affinity with democracy. Extending political power to nonpropertied classes poses a potential threat to the concentration of economic wealth. It is instead "the growth of a counter-hegemony of subordinate classes and especially the working class—developed and sustained by the organization and growth of trade unions, working-class parties and similar groups that is critical for the promotion of democracy."[64]

The second analytical cluster Rueschemeyer, Stephens, and Stephens draw attention to is the balance between civil society and the state, which in the developing world is critically shaped by the specific institutional legacies of colonial rule. A balance that effectively limits the state's repressive or co-optive capacity is favorable to the emergence of prodemocratic forces. Taken alone, this argument could readily be confused with the institutional determinism of conventional interpretations of the roots of Indian democracy. Indian democracy in this perspective is largely viewed as an extension of the bureaucratic structures, rule of constitutional law, and representative institutions bequeathed by the British.[65] That other British colonies proved much less propitious for democracy—the most notable of which is of course Pakistan—underscores the inadequacy of this explanation. Existing institutional configurations certainly do matter, but not in their own right. Instead, the balance between state and civil society—as Rueschemeyer, Stephens, and Stephens argue—becomes critical to the prospects of democratization only insofar as it affects the balance of class forces. Specifically, the density and robustness of civil society is critical on two counts: it creates spaces and organizations through which subordinate groups can mobi-

[63] Ibid., 8.

[64] Ibid., 50.

[65] See Myron Weiner, *The Indian Paradox: Essays in Indian Politics*, ed. Ashutosh Varshney (New Delhi: Sage Publications, 1989). Alternatively, Jalal has argued that there is in fact little difference between Pakistan and India in that both have inherited state forms that she [illegible] authoritarianism": Ayesha [illegible]

lize independently of the influence of dominant groups, and its curtails the repressive capacity of the state.

How then can Moore and Rueschemeyer, Stephens, and Stephens shed light on our understanding of the degrees of democracy across India? The consolidation of democracy in India combined elements of both a favorable institutional context and a successful, but limited mobilization of subordinate groups. The rise of the Indian National Congress as the agent of democratic transformation was clearly facilitated by the truncated but not insignificant rights of association and representation afforded by British colonial rule. Up until the 1920s the INC was very much a creature of colonial liberalism—a reformist party of urban professionals and progressive elements of the economic elite. It was only with the advent of Gandhi that the Congress struck roots in the countryside and was transformed into a legitimate national organization with a mass base. As Moore argues, this class coalition of peasants and urban elites bolstered the political power of what was otherwise a weak bourgeoisie and marginalized the colonial class of feudal landlords.[66] Because Gandhi's Congress did shift the balance of power against landlords and did draw the masses into the political arena, it put India firmly on the road to democracy.

But the "peasant masses" that were mobilized as part of the anticolonial coalition were in fact deeply divided along class and caste lines. Agrarian tensions (which had given rise to a wave of peasant rebellions in the early part of the century) were subordinated to the nationalist logic of class conciliation.[67] Subordinate classes, specifically landless laborers, tenants, and poor peasants, whose class position generally coincided with untouchable or lower-caste status, were as such never mobilized on their own terms as an independent political force with interests and strategies of their own. Thus, although India's transition to democracy was impelled by mass mobilizations, in the final analysis it represented an elite-dominated pact that was subsequently institutionalized by the electoral domination of Congress. During the Nehru period Congress secured its position by mobilizing a wide range of interests through an elaborate network of local notables (mostly dominant landed castes) who controlled and delivered votebanks in exchange for state patronage.[68] In this manner Congress—and hence

[66] Moore (fn. 11), 354–77.

[67] Gyan Pandey, "Peasant Revolt and Indian Nationalism: The Peasant Movement in Awadh, 1921–22," in Ranajit Guha and Gayatri Spivak, eds., *Selected Subaltern Studies* (New York: Oxford University Press, 1988).

[68] For a seminal analysis ...

democracy—embedded itself in rural areas (winning five consecutive and openly contested national elections) without challenging existing forms of social domination. There was in effect no rupture with the ancien régime, which, as Moore argues, has historically been the precondition for the successful consolidation of democracy. Beyond removing the larger parasitic landlords, land reforms were carried out indifferently, and though the size of the propertied class expanded, land ownership remained largely confined to local dominant castes. Efforts to develop and democratize local institutions for community development and democratization, such as village councils and cooperatives, were defeated by elite capture. In sum, the state enmeshed itself in a matrix of accommodations and patronage networks and thus undermined its ability to pursue transformative projects, including the extension of public legality to rural areas. While this mode of engagement of society did provide a basic framework for political order, it failed to build institutionally robust arenas of civic associationalism and severely curtailed both the instrumental and the authoritative efficacy of the state. To borrow a phrase from Gramsci, the state-cum-Congress could rule, but t could not lead.

Democracy was born under very different conditions in Kerala and ıas accordingly traveled a very different path. In understanding the diergence from the national pattern, one has to begin with Kerala's his- ɔry of social mobilization and specifically the early predominance of lass mobilization. In the first part of this century social mobilization ı Kerala coalesced around three distinct axes. A social reform moveent led by caste-based associations directly challenged the social and stitutional inequities of the caste system but stopped short of attackg its socioeconomic base. The nationalist movement, led initially by e Congress, followed the class conciliatory pattern of the national ›vement. Finally, there were significant but mostly inchoate instances agrarian rebellion, most notably the Mapilla uprising of 1921, as well some grassroots efforts to organize support for land reform.

The turning point came in the 1930s. Disillusioned with the accomdationist line of the National Congress, socialist elements within the ala Congress began to link—both organizationally and ideologiy—the nationalist and social reform movements to agrarian discons. Within a decade the struggle against British rule and the ›cratic princely states of south Kerala had become a broad b- r-class movement [illegible]

the CPI in 1941—played a catalytic role, forming a politically unified class movement from what were otherwise geographically and economically very different groups. In the northern part of the state, Malabar, agrarian conflict pitted a class of rack-rented tenants, many of whom were Muslim, against a powerful class of upper-caste Hindu landlords. In Travancore and Cochin incipient class mobilization had taken the form of trade unionism among both factory workers in British coir factories and lower-caste laborers in the rice fields of Kuttanad.

Classes are not given structurally. They are, in Przeworski's succinct formulation, "formed in the course of struggles."[69] Those struggles take shape on multiple fronts and affect the process of class formation only inasmuch as they come to define new identities. Building on existing repertoires of contention (to use Tilly's term),[70] the communists wove together the themes of social dignity and justice of the caste reform movement, the demands for economic redress of the agrarian movement, and the democratic aspirations of the nationalist movement into a coherent and sustained ideological attack against colonialism and the feudal class/caste structure. In portraying excessive rents, labor servitude, wage exploitation, caste indignities, poverty, and colonial subordination as part of the same of system of domination, the communists recast a world of complex overlapping social positions and identities into "a new world populated only by the working masses and the exploiting classes."[71]

The CPI's early successes were more those of a social movement than those of a political party. Internally, the success of social movements relies on the capacity for creating what Tarrow calls "collective action frames."[72] "Ideology," as David Apter writes, "dignifies discontent, identifies a target for grievances and forms an umbrella over the discrete grievances of overlapping groups."[73] Driven more by events and concrete struggles than by theoretical insight,[74] the party actively translated the fragments of subaltern identities articulated in the social justice discourses of existing movements into a cohesive class agenda of

[69] Adam Przeworski, *Capitalism and Social Democracy* (Cambridge: Cambridge University Press, 1985), 69.

[70] Charles Tilly, *From Mobilization to Revolution* (Reading, Mass.: Addison-Wesley, 1978).

[71] Dilip Menon, *Caste, Nationalism and Communism in South India: Malabar 1900–1948* (New Delhi: Cambridge University Press, 1994), 146.

[illegible] *in Movement* (Cambridge: Cambridge University Press, 1994), chap. 7.

sociopolitical transformation. It was Leninist party discipline and mas organizations that leveraged movement demands, but as Menon has sc convincingly shown, it was the party's cultural activities that exposec the inequities and indignities of the traditional social order.[75] Templ festivals became venues for political meetings. Theatrical troops tourec the countryside, presenting plays that popularized Marxist ideas, exposed caste injustices, and celebrated revolutionary heroism. Village-level "reading rooms" became incubators for a new secular culture tha transcended caste and religion. A new literary movement producec popular socialist novels, and the party newspaper, *Prabhatham* launched in 1936, provided news of union activities, peasant struggles and factory conditions. The party's active role in the revival of Malayalee culture was underscored politically by its championing of the caus of a United Kerala, a position it had first embraced in 1942, when th Congress still refused to interfere in the "internal affairs" of the princel states. In sum, the communists built a broad-based following througl an integrative strategy of sponsoring mass organizations and by entrenching itself in civil society. In a pattern reminiscent of West European mass parties, the rise of secondary associations—the buildin blocks of civic capacity in Putnam's model—thus had distinctly political origins in Kerala.[76]

The strategic success with which the Communist Party situated itself at the confluence of social movement and structure is only part o the story. A key enabling factor was that the communists were operating within a favorable and really quite unique institutional and politica environment. Historically, agrarian communism has been a violent affair, marked by either revolutionary violence or state violence or both eventuating in most cases in authoritarian regimes. The CPI did not entirely eschew insurrectionary methods, and the state did not entirely resist resorting to violent repression. But in comparative terms the rise o communism in Kerala was a rather peaceful process, one that took plac largely in the trenches of a colonial political order that had graduall conceded limited rights of association and opposition to its colonia subjects and whose most repressive reflexes were constrained by its ow liberal pretensions. While subject to occasional censorship, the com-

[75] Menon (fn. 71), chap. 5.

[76] In a critique of Putnam's cultural and historical explanation of the sources of civicness in Nort Italy, Tarrow notes that in the past and the present "electorates were deliberately mobilized on the basi of networks of mass organizations and social and recreational associations; and in both, civic competence was deliberately developed after World War II as a symbol of the left-wing parties' governin

munists operated for the most part in public spaces, building a d network of unions, farmer associations, schools, libraries, cult organizations, and press organs. Their protest activities ranged from Gandhian tactics of sit-ins, fasts, boycotts, and civil disobedience those more typically associated with working-class movements strikes, pickets, and marches. If a comparatively favorable (by t standards of colonial states) balance between civil society and the sta extended the repertoire of contention, it also made possible th "ratchet-effect strategy" of incremental gains that sustained clas mobilization.[77]

With independence, the state-society balance was further tipped in favor of mobilization. A formally democratic state presented an especially attractive object and venue of mobilization, and the process of class formation became inextricably linked with the process of state building. In Malabar the party pushed for land reforms and won the first district elections. In Travancore it organized large-scale strikes of agricultural and industrial workers, demanding state intervention in enforcing minimum wages and regulating work conditions. In 1956 the state was unified along linguistic lines and the following year the CPI captured a majority of seats in the legislative election, becoming the first democratically elected communist government in the world. Though short-lived (the government was illegally deposed by the center in 1959) the ministry represented a threshold in the trajectory of Kerala's mass-based democracy. It marked the ascendancy, achieved through the ballot box, of the poor and propertyless, social groups that in a few short decades had gone from complete social and economic subordination to political power. The 1957 government set into motion a series of reforms that over the next two decades would transform the face of Kerala's agrarian social structure. It also set a standard for state intervention and social welfare from which no subsequent government has strayed.[78] Most importantly, the political tide on the agrarian ques-

[77] Ronald Herring, "From Fanaticism to Power: Ratchet Politics and Peasant Mobilization in South India, 1836–1956" (Paper presented at the annual meeting of the Association of Asian Studies, Honolulu, April 11–14, 1996).

[78] The demand-side dynamic for health and education in Kerala predates the democratic period and can be tied to the comparatively progressive educational and public health policies of the princely state of Travancore. See Gita Sen, "Social Needs and Public Accountability," in Marc Wuyts, Maureen Mackintosh, and Tom Hewitt, *Development Policy and Public Action* (Oxford: Oxford University Press, the Open University Press, 1992). The transition from the Brahmanical paternalism of Travancore, which targeted communities (which in many ways has its contemporary counterpart in the populism of the Indian state) to the social citizenship of Kerala's [illegible]

ad turned. The debate on land reform took center stage for the two decades with all political forces eventually aligning themselves vor of reform.

Mobilization and Democratic Deepening

e birth of democracy in Kerala was indelibly branded with the logic social transformation. Well into the 1970s, the politics of class strug- occupied center stage. Throughout this period, the conflicts between tenants and landlords, labor and capital, upper caste and lower caste were acrimonious. The communist agenda was one of radical transformation, its methods those of large-scale agitation and labor militancy. Taking its cue from communist organizational successes, the Congress built its own mass organizations. Politics became synonymous with popular mobilization and Kerala often appeared to be teetering on the brink of ungovernability, with hypermobilization threatening to overload political institutions.

Nevertheless, because class conflicts had evolved within a framework of democratic rules of the game that were well established, enjoyed a high degree of popular legitimacy, and were accepted by all the principal players, they did not result in breakdown or disintegration. Excesses on either side were checked, moreover, by the subnational character of the playing field. Substantively modest but symbolically important successes on the parliamentary front had made "bourgeois democracy" and a reformist line acceptable pragmatically, if not ideologically, to the communists. The weakness of the CPI at the national level ruled out revolutionary tactics. Right-wing mobilization, which at times flirted with authoritarian reaction, was curbed by the commitment of national elites to electoral democracy. Finally, while agrarian communism did produce comprehensive land reform (1970) and agrarian labor legislation (1974) that virtually eradicated the material and social power of landed elites, the process was slowed and defused by constitutional procedures, guarantees of private property, and drawn-out political negotiations. Large-scale peasant mobilizations in the 1960s and 1970s provided the impetus for agrarian structural transformation. But just as significant was the process through which these reforms wended their way through public debates, courts, legislative enactments, and the bu-

was dramatically closed in the post-1957 period. See T. N. Krishnan and M. Kabir, "Social Intermedi-

reaucracy, firmly embedding the state in society. As such, class strugg
could evolve and develop within boundaries that more or less preclude
the resort to revolutionary violence or elite repression.

In keeping with the Rueschemeyer, Stephens, and Stephens analyt
ical framework, then, this account of democratization points to th
centrality of subordinate actors and highlights the historically contin
gent and institutionally bounded circumstances of class formation. Th
centrality of the *process of class formation* rather than the structural char
acter of the class actor is underscored by the fact that the historical pro
tagonist was not an industrial working class. Instead it was
broad-based and loosely configured agrarian class forged from the his
torical convergence of social movements, the structural congruence o
social domination (caste) and economic exploitation (class), and th
tactical successes of a communist party born at the intersection o
agrarian radicalism and parliamentary politics.

The historic conjunction of social forces and political organizatio
provided the critical wedge that pried open the ancien régime an
opened up new possibilities for democratic politics. If this foundation
moment transformed the playing field by shifting the balance of powe
it did not inexorably set Kerala down the path of democratizatio
Given that class interests and alignments constantly shift and mus
compete with other bases of mobilization, the sustained effectiveness o
subordinate class politics has to be explained. The temptation here is t
follow Kohli's seminal work on the CPM in West Bengal and argue th
a programmatic and disciplined political party has played the critic
role in aggregating and sustaining lower-class interests.[79] Much as i
the case of the West Bengal CPM, an ideologically cohesive part
governed internally by "democratic centralism" has kept factionalis
in check, institutionalized lower-class interests, and increased th
effectiveness of government policies, especially in the area of povert
alleviation.

A party-centered argument that emphasizes the centrality of organ
izational capacity in channeling mobilizational dynamics provides onl
one side of the equation, however. In contrast to its twenty-three con
secutive years in power in West Bengal, the CPM in Kerala has ruled onl
intermittently and never for even two consecutive terms. The Keral
CPM's critical role has been less a function of its governance capacit
than of its mobilizational capacity. Having found itself periodically i
the opposition, the CPM has retained much of the social movement dy

[79] Kohli (fn. 7, 1990), chap. 10.

namic from which it was born by having to continually reinvigorate its mobilizational base and reinvent its political agenda. As a *social movement party*, the communists have thus busied themselves with the task of occupying the trenches of civil society, building mass-based organizations, ratcheting up demands, and cultivating a noisy but effective politics of contention. This has provided a continuous presence and effectiveness for subordinate groups even when out of power.[80] And much as the party has helped sustain movement politics, conversely, its immersion in civil society has kept oligarichal tendencies in check and has allowed for an uncommon degree of political learning, as witnessed by the party's recent embrace of the "new" social movement project of grassroots empowerment (see discussion of decentralization campaign below).

Adding to the general assertiveness of civil society in Kerala has been the bandwagon effect of other political parties embracing mass-mobilizational politics. With two equally balanced, if not equally effective, political fronts actively nurturing and courting support from civic associations, political parties in Kerala have had a crowding-in effect on civil society. Nowhere in India have the contentious repertoires of social movements become such an intrinsic part of routine politics. In sum, though there has been little regime continuity in Kerala, there has been continuity of political participation and access. Effective democratic governance has in other words had less to do with the institutional character of the political party system than with the dynamic interaction of political and civil society.[81]

THE AFFINITIES OF CLASS AND DEMOCRACY

If there is, as I have argued, a positive relationship between the articulation of class demands and the deepening of democracy in Kerala, how do we come to terms with this theoretically, especially in light of assertions by some transition theorists that the conflictual nature of class-based politics is inherently destabilizing?[82] In ideal-typical terms,

[80] In contrast, the CPM in West Bengal, having enjoyed the fruits and resources of state power for twenty-three years, has become far more bureaucratic and clientelistic. See John Echeverri-Gent, "Public Participation and Poverty Alleviation: The Experience of Reform Communists in India's West Bengal," *World Development* 29 (October 1992).

[81] Linz and Stepan (fn. 1) write that "properly understood, democracy is more than a regime; it is an interacting system" (p. 13). A key dynamic in the system derives from the distinct but complementary roles that political and civil society play in shaping and representing values and interests. See Linz and Stepan (fn. 1), 8–9.

[82] To quote Huntington: "Democratic regimes *that last* have seldom, if ever, been instituted by mass popular action." Samuel Huntington, "Will More Countries Become Democratic?" *Political Science Quarterly* 99, no. 2 (1984), 212, emphasis added.

[illegible] binds state autl [illegible] other. The glue of that relationship is both and normative: utilitarian in the sense that states provide pu in exchange for compliance and resources from their citizen tive in the sense that the strength of the relationship (the qu scope of public goods, the degree of citizen compliance) is born which in complex and differentiated societies can be secur through processes of consultation that are sufficiently inclusive a tained to be broadly perceived as fair. As Tilly has recently obser

> *In the course of democratization, the bulk of a government's subject populati acquires binding, protected, relatively equal claims on a government's agents, a tivities and resources. In a related process, categorical inequality declines i those areas of social life that either constitute or immediately support participa tion in public politics. Finally, a significant shift occurs in the locus of interper sonal networks on which people rely when undertaking risky long-term enterprises such as marriage, long-distance trade, membership in crafts,* and investment of savings; such networks move from evasion of governmental detection and control to involvement of government agents and presumption that such agents will meet their long-term commitments. Only where the three sets of changes intersect does effective, durable democracy emerge.[83]

To understand why these processes became reinforcing in Kerala, th analytical key lies in carefully untangling the relationship between stat intervention and actual patterns of demand aggregation. Subordinat class movements make encompassing demands, usually framed by call for social leveling and protection from the injustices of purely market-based resource distributions.[84] In a poor and deeply hierarchical societ such demands are in effect public goods and lend themselves to the instrumentalities of the modern, bureaucratic state. In contrast, factional or community-based movements make particularistic demands that tend to be mutually exclusive. While the form of such demands can be democratically managed, the substance is secured through clientelism, which has well-known corrosive effects on state capacity and by definition compromises the associational autonomy of subordinate groups.[85]

The demands that emanated from class mobilization in Kerala fall into three broad categories. The first was for land reform, which sought not only to redistribute property but, just as importantly, to eradicate the social and economic basis of landlordism. The second was for the

[83] Charles Tilly, "Processes and Mechanisms of Democratization," *Sociological Theory* 18 (March 2000), 8.

[84] Joshua Cohen and Joel Rogers, "Secondary Associations and Democratic Governance," *Politics and Society* 20, no. 4 (1992).

[85] See Fox (fn. 18).

and protection of the rights of wageworkers. This aimed the forms of labor-tying that prevailed in agriculture and r workers with collective-bargaining rights. The third was ension of basic social services, especially health and educa- nands that in fact predated the CPI and enjoyed widespread across all classes. Now in their intrinsic character, all these de- amounted to expanding the scope and the prerogatives of the e state. The practical effect of this was to extend the reach of t legality into arenas (for example, caste relations, the informal r) that in the rest of the country continue to be governed by ex- emocratic authority.

Class mobilization constituted citizens in Kerala. Freed of the ideo- gical and social dependency on corporate groups and powerful pa- ons, horizontally organized subordinate groups could proactively ssociate and openly articulate demands. The high level of demand naking, itself facilitated by a favorable balance of state and civil soci- ty, drew the state in. With each successful intervention, the authority f the state and formally recognized interest groups displaced the au- hority of traditional power brokers, thus slowly but surely chipping way at the ties of social and economic dependency that had character- zed the preindependence social order. A welfare state with its statutory ntitlements replaced a caste-based moral economy with its discrete nd asymmetrical reciprocities. Formal contracts and monied wages ook the place of attached labor and payments in kind. A division of abor rooted in a highly stratified and rigid caste system was subjected o market dynamics bounded by the state-sanctioned bargaining capac- y of associated workers. Work relations embedded in the extraeco- omic power of landlords and merchant bosses were displaced, albeit nevenly, by labor legislation, labor inspectors, and formal grievance rocedures. Access to education and health was made a function of cit- zenship rather than of social position. And increasingly, conflict reso- ution took the form of collective bargaining, tripartite consultations, nd judicial review. In Tilly's terms,[86] trust in Kerala has shifted from nterpersonal networks rooted in categorical social inequalities to in- esting in state institutions, an investment informed by iterated experi- nces with the state and a resulting faith that state authorities will fulfill heir commitments.

Many observers of Kerala have argued that labor militancy, state in- ervention, and increased social expenditures have exacted a high price

[86] Tilly (fn. 83).

on economic growth.[87] The 1975–85 decade in particular was mark by economic stagnation in both industry and agriculture. The extent which redistributive reforms directly contributed to this crisis is dif cult to establish, especially given the importance of other factors (a h torically weak industrial base, resource scarcity and high populati density, the absence of a local entrepreneurial class). I have treated tl question at length elsewhere, but three points can be made.[88] First, ev if there is indeed a zero-sum trade-off between equity and growth, o would be hard pressed to find another Indian state where this trade-c has been managed more effectively. Despite low levels of growt poverty in Kerala has fallen faster than in any other state and it no boasts the most extensive safety net in the country.[89] The high levels basic and technical education and a well-developed public infrastru ture represent critical assets for future growth. Moreover, as the mark economy in India expands, there will necessarily be increased demar for primary education, access to health care, and social protection—a sectors in which Kerala has already made significant infrastructural, ac ministrative, and fiscal investments—and Kerala will find itself at a sig nificant comparative advantage. Second, to view the equity-growt trade-off in narrowly economic terms obscures the important politic changes that have taken place. As noted earlier, the CPM and its unior responded to the economic crisis by abandoning wage militancy an embracing a strategy of class compromise that has focused on ir stitutionalizing industrial conflict and increasing worker productivit The state has also become much more aggressive in attracting investo and nurturing key growth sectors. How effective these political an policy shifts have been is difficult to determine (especially since Ke ala is a subnational state with limited macroeconomic powers), bu Kerala's economy has experienced a significant turnaround since th mid-1980s.[90]

[87] See Joseph Tharamangalam, "The Perils of Development without Economic Growth: The D velopment Debacle of Kerala, India," *Bulletin of Concerned Asian Scholars* 30, no. 1 (1998). For r sponses to Tharamangalam, see *Bulletin of Concerned Asian Scholars* 30, no. 3 (1998).

[88] Heller (fn. 48, 1999).

[89] A recent World Bank study found that from 1957–58 to 1990–91 Kerala experienced the mc rapid decline in poverty of any major state, including the Punjab and Haryana, India's capitalist grow success stories. Guarav Datt and Martin Ravallion, "Why Have Some Indian States Done Better Th Others at Reducing Poverty?" Policy Research Working Paper no. 1594 (Washington, D.C.: Wor Bank, 1996).

[90] Between 1986–87 and 1993–94 the factory sector grew at an annual average of 9.8 percent, w above the national average of 5.6 percent, and agriculture grew at 5.5 percent. A recent study on ne investments in Kerala found a "tremendous increase" since 1991–92. Sunil Mani, "Economic Libera isation and Kerala's Industrial Sector: An Assessment of Investment Opportunities," *Economic and P litical Weekly* (August 24–31, 1996), 2326.

Finally, if the encompassing logic of subordinate class politics has :en central to Kerala's democratic trajectory, it has hardly been im- une to factionalizing and rent-seeking tendencies. As the role of the ate in regulating economic life and distributing resources has grown, have distributional coalitions. Moreover, the solidaristic politics of e left have become increasingly difficult to sustain. The communists ave never represented a fundamental or essential class. Classes are po- tically constituted and as fluid entities must constantly be reinvented. urther, the politics of class in Kerala has been played out against the ıckdrop of a rapidly changing social structure that has generated new eavages and new alliances. The embourgeoisement of the poor peas- ıtry following the transfer of property rights from landlords to ten- ıts in the 1970 land reform eroded the structural basis of agrarian ommunism.[91] Sluggish industrial growth has limited the size of the ıdustrial workforce. And the increase in the size of Kerala's middle ass, fueled by the expansion of the welfare state and remittances from erala's huge out-of-state labor force, has further weakened the tradi- onal class base of the left.

The political, institutional, and structural legacies of class mobiliza- on in Kerala continue, however, to provide a strong foundation for emocratic development. The erosion of the social and economic ower of landed elites has weakened the expression of a range of frag- ıentary and parochial interests that continue to dominate national olitics. A dense network of intermediate organizations and institu- ons provides multiple points of interface between autonomous asso- ations and state agencies. Patterns of state-society engagements have roduced an informed and engaged citizenry. Welfare entitlements, age legislation, market regulation, and other forms of social protec- on have not only substantially insulated wage earners from the more omizing effects of market forces but have also secured a considerable egree of institutionalized bargaining capacity for large segments of the age-earning classes. And the provision of a wide range of public oods has underwritten a baseline solidarity, including middle-class ıpport for the welfare state.

Despite the decline of the structural significance of class—at least as was historically constituted in Kerala's agrarian transition—the CPM d its allies have been able to fashion a new politics of popular devel- ment, one that has been specifically articulated around a project of

[91] See Ronald Herring, "From Structural Conflict to Agrarian Stalemate: Agrarian Reforms in uth India," *Journal of Asian and African Studies* 26, no. 3–4 (1991).

further democratization and that has broadened the developme agenda beyond redistributive issues to include questions such as the vironment, administrative and political decentralization, local resou planning, and the reinvigoration of an overly bureaucratized coope tive movement.[92] Concretely, the CPM-led government that returned power in 1996 launched the People's Campaign for Decentraliz Planning, an initiative widely recognized to be India's most ambitio effort at comprehensive decentralization.[93] As part of an open atta against rent-seeking fiefdoms within the state and the top-down log of bureaucratic planning, substantial financial and administrative re sources have been devolved to local-level governments and broad-base participation in local decision making has been mobilized. If the polit ical opening for decentralization was orchestrated by the CPM fro above, it is civil society that provided the critical ideological and mobi lizational resources for the campaign. The campaign's discourse of au tonomy, local initiative, transparency, sustainability, and accountabilit is the language of social movements, not of technocrats or Leninists Most of the techniques and favored projects of the campaign com from a repertoire of practices that NGOs and proactive local govern- ments have been developing for years. The more than one hundred thousand volunteers who have been trained to provide organizational and technical assistance to local governments have been recruited from civil society and not from the party's traditional mass organizations. Most critically, it is through overlapping membership ties between the CPM and the independent, grassroots KSSP that CPM reformers could experiment with ideas outside the somewhat doctrinaire straitjacket of the party itself and build political support for a strategy of mobilization that reaches beyond the party's traditional base of support. While it is too early to judge the sustainability of this decentralization initiative, the very existence of a political project specifically centered on promot- ing new forms of democratic participation could not be more telling, given the increasing involution and divisiveness of national politics.[94]

[92] See T. M. Thomas Isaac and E. M. Shridharan, "Kerala's Development and Its Politics," *Marxist Samvadan* (Trivandrum) (October–December 1992); and Olle Tornquist, *The Next Left? Democratisation and Attempts to Renew the Radical Political Development Project* (Copenhagen: NIAS reports, no. 24, 1995).

[93] See Richard Franke and Barbara Chasin, "Power to the Malayalee People," *Economic and Political Weekly* 31, no. 10 (1997).

[94] These observations are based on research conducted in summer 1997 and 1999.

CONCLUSION

Under the conditions of late development, the challenge of democratic deepening appears to be even greater than the challenge of democratic transitions. Weak states, severe economic inequalities, and the resilience of predemocratic sources of authority have made it difficult to translate the political opportunities afforded by democratic institutions into the effective exercise of citizenship rights and substantive gains. For some authors, these obstacles appear to be insurmountable. Weyland argues that efforts by state reformers in Brazil to pursue equity-enhancing reforms have been repeatedly frustrated by entrenched oligarchical interests and the pervasive organizational fragmentation of Brazilian politics.[95] In South Africa economic concessions made to the white minority and to market forces and the authoritarian legacies of indirect rule under apartheid have frustrated the promise of rapid democratization.[96]

Yet if we peer below the national level, it is possible to find islands of democratization. In a number of Brazilian municipalities, an alliance of the Partido dos Trabalhadores (a social movement party) and civic groups has broken the hold of oligarchical elites by introducing new institutions and processes of popular participation in municipal policymaking that have had measurable redistributive effects.[97] In South Africa the embrace at the national level of conservative promarket policies and the centralizing tendencies of the ANC must be contrasted with the vitality of NGOs and the strength of proredistributive alliances in the country's largest metropolitan areas. The point here is that while national-level conditions in much of the developing world may not be favorable to democratization, we need to pay more attention to how the subnational reconfiguration of political forces can transform state-society relations and produce democracy-enhancing effects. Democracy, in other words, can be built from the bottom up.

The case of Kerala here is especially instructive. In a pattern that bears a strong resemblance to European social democracies, the procedural, effective, and substantive dimensions of democracy have become mutually reinforcing. That this dynamic was played out against a back-

[95] Weyland (fn. 19).

[96] For an analysis of economic concessions made during the negotiated transition, see Hein Marais, *South Africa: Limits to Change* (London: Zed Books, 1998). Mamdani (fn. 30) argues that "decentralized despotism"—the entrenched authority of chiefs—has stunted the development of civil society in postapartheid South Africa.

[97] See Rebeca Abers, "From Ideas to Practice: The Partido dos Trabalhadores and Participatory Governance in Brazil," *Latin American Perspectives* 23, no. 4 (1996).

changes and very low levels of
rather than succeeded, capitalist
independent effect of political proces
the critical role of subordinate classes. The
was the existence of a procedurally robust demo
provided critical spaces in which subordinate groups could
But if Kerala has parted from the national pattern, the cause
located in cycles of state-society interactions that were trig
class-based politics and produced three broad effects.

First, in a society marked by profound social and econo
equities, the forging of a lower-class movement into a cohesive o
zational force decisively shifted the balance of power in fa
subordinate groups, paving the way for redistributive reforms. The
forms in turn eroded the economic and social power of landed
and strengthened the associational autonomy of lower classes.

Second, the logic of class mobilization—redistributive conflic
drew the state in and created the political impetus for getting the de
ocratic state to do what it does best, that is, provide public goods. M
visible was the universal provision of basic services, as well as instit
tional reform. Less tangibly, but just as crucially, the intensity and sus
tained character of economic conflicts necessitated lasting and
routinized state interventions, rather than payoffs or selective co-optation. The result has been the creation of a rich fabric of mediating institutions governed by legality and democratic authority. The state's demonstrated capacity to effectively mediate distributional conflicts has strengthened the legitimacy of democratic institutions. Another effect of state intervention has been the consolidation of a critical, but much neglected attribute of any robust democracy—a socially regulated market economy.[98] The affinity between democracy and markets is not functionally given, it is historically constructed. The bourgeoisie has historically supported democracy because it provides for the accountability, third-party arbitration, and rule-bound enforcement of laws and contracts that market economies need. Subordinate classes support democratic institutions because they are the means to securing a more equitable distribution of wealth and some degree of protection from economic downswings. Kerala remains a poor economy, but the extent

[98] This follows Karl Polanyi, *The Great Transformation* (Boston: Beacon Press, 1944). See also Linz and Stepan (fn. 1), who write about a set of "socio-politically crafted and socio-politically accepted norms, institutions and regulations, which we call *economic society*, that mediate between state and market" (p. 11).

ments of the informal economy have been subjected
d social protection stands in sharp contrast to the
ations and social vulnerabilities that characterize most
nomy. Because socially regulated capitalism benefits
nts of the population, is it far more conducive to democ-
n is laissez-faire capitalism.

e logic of class politics has strengthened civil society. It has
t through the small group dynamic of trust and reciprocity
ed by many civil society theorists, but rather through the
e of broader solidarities that were forged from a history of
On the one hand, repeated cycles of mobilization have cre-
anizations and networks that cut across traditional social cleav-
us broadening the associational scope and quality of public life.
al movement theorists have argued, participation in movements
ositive spillover effects for democracy in that it creates new soli-
ies and nurtures a culture of civic engagement.[100] On the other
, class-based mobilization has created forms of conflict that lend
nselves to compromise and encompassing solutions. Unlike many
r forms of claim making, pursuing redistributive demands in a cap-
ist economy (in which future growth and employment depend on
vate investment) reveals the interdependence of class interests.[101]
hesive labor movements in a private property economy can and do
t strategically,[102] and the resulting compromises tend to emphasize
operative and inclusionary social policies.

ovides an excellent critique of the overemphasis of the civil society literature on com-
nity-based origins of civic behavior and the resulting neglect of the role that more aggregated forms
ssociation that transcend local community (social movements, advocacy groups) can play in nur-
ing civic identities. Debra Minkoff, "Producing Social Capital: National Social Movements and
il Soceity," *American Behavioral Scientist* 40, no. 5 (1997).

[100] Ibid., 612.

[101] Claus Offe, "Homogeneity and Constitutional Democracy: Political Group Rights as an Answer
Identity Conflicts?" (Paper presented at the Conference for the Study of Political Thought, April
6, Columbia University, 1997), 6.

[102] See Adler and Webster (fn. 12).

THE IDEOLOGICAL DETERMINANTS OF LIBERAL ECONOMIC REFORM
The Case of Privatization

By HILARY APPEL*

FOLLOWING the collapse of the communist regimes in Eastern Europe, one of the first priorities of the reform governments was to transform the system of ownership. The new governments overwhelmingly chose to pursue a broad program of sweeping privatization, transferring the bulk of state property to the private sector. Given that most property was controlled and owned by the state and that there was no historical precedent for privatization on an economy-wide scale, policymakers had to design from scratch a program to transfer the wealth of the state to private hands. With few exceptions, East European privatization officials appointed to this task concluded that it would be too slow and unwieldy to rely on the piecemeal sale of enterprises to private investors, as was typical in West European privatization. More commonly, they developed programs to transfer the ownership of state enterprises to private individuals and groups through mass distribution programs, to be followed by or supplemented with piecemeal sales.

As one can imagine, these policymakers and technocrats could design the mass privatization program in ways that held different distributional consequences for society, such that some groups would benefit more than others, for example, citizens over foreigners, one former elite group over another, managers over labor, and so on. In the Czech Republic the government designed a mass privatization program in which there were no special privileges for any particular groups in society.

* The author wishes to express her gratitude for the helpful comments and suggestions on earlier drafts from Andrew Barnes, Thomas Callaghy, Margaret Hanson, Meredith El Nems, Steven Fish, Denise Powers, and Peter Rutland. In addition, the author wishes to acknowledge the Woodrow Wilson Center and the Kennan Institute for Advanced Russian Studies for their generous support. The opinions expressed are those of the author and do not necessarily reflect the institutions or individuals acknowledged above. Responsibility for any errors is assumed entirely by the author.

Once the state transformed medium-size and large firms into corporations, any adult citizen could obtain from the government vouchers that could be exchanged for shares in the newly created corporations. There were no explicit privileges or discounts for participants in the voucher program, although foreigners could not participate. In Albania and Hungary the privatization programs similarly avoided granting privileges to particular domestic economic and social groups; Hungarian privatization is unusual, moreover, for encouraging foreign participation over mass participation.[1]

In Russia privatization officials initially adopted a mass, voucher-based program modeled closely on the Czech experience. There was, however, one crucial difference that became more pronounced over time. Although the Russian program started out distributing vouchers equally to all citizens, the program later added enormous privileges for employees. In fact, in 72.5 percent of Russian enterprises privatized through the program, 51 percent of shares were reserved for the employees of that enterprise.[2] Privatization officials in other states similarly designed programs to benefit workers. In Bulgaria 20 percent of shares were offered at 50 percent discount; in Lithuania 50 percent of shares were available to managers and workers at a discount; and in Ukraine a leasing program essentially transferred for free large blocs of shares to worker cooperatives.[3]

In order to explain the distributive benefits of mass privatization programs across Eastern Europe, the academic literature relies primarily on material interests and power as explanatory variables, paying little if any attention to ideological and cognitive factors. The goal of this

[1] While no other postcommunist privatization program was as neutral as the Czech program regarding employee ownership rights, Albania's program is a close second. The particularist benefits for enterprise employees were limited to special bidding arrangements that were poorly executed and with minimal effect. On Albanian privatization, see Igor Artemiev and Gary Fine, "Albania," in Ira Liberman et al., eds., *Between State and Market: Mass Privatization in Transition Economies* (Washington, D.C.: World Bank, 1997), 177–80. On Czech privatization, see Michal Mejstrik, *The Privatization Process in East-Central Europe: Evolutionary Process of Czech Privatizations* (Boston: Kluwer Academic Publishers, 1997).

[2] While estimates vary between 70 and 77 percent depending on the source, 72.5 percent was cited by the former chairman of the State Property Committee in Alfred Kokh, *The Selling of the Soviet Empire* (New York: Liberty Publishing House, 1998), 146.

[3] For an extensive discussion of Polish privatization, see Mitchell Orenstein, *Out of the Red: Building Capitalism and Democracy in Post-Communist Europe* (Ann Arbor: University of Michigan Press, forthcoming). On the distribution of benefits in Russian mass privatization, see Peter Rutland, "Privatisation in Russia: One Step Forward Two Steps Back?" *Europe-Asia Studies* 46 (November 1994); Hilary Appel, "Voucher Privatization in Russia: Structural Consequences and Mass Response in the Second Period of Reform," *Europe-Asia Studies* 49 (December 1997); and A. Radygin, *Reforma sobstvennosti v Rossii, Na puti iz proshlogo v budeshchee* (Property reform in Russia: On the way from the past to the future) (Moscow: Respublika, 1994). For a comparison of fifteen mass privatization programs, see Liberman (fn. 1), 10–13, Table 2.

article is to make explicit the mechanisms by which ideology influences the design of privatization programs. The next section discusses why existing power-based and interest-based approaches to the study of privatization and property rights change must be expanded to take non-material, ideological factors into account. The third section presents a four-tiered framework of how ideas and ideology determine the desigr and development of new property rights systems. The final section discusses the larger theoretical implications of incorporating an ideological component into theories of political economy.

Standard Approaches to Property Rights Change

The variation in the designs of privatization policies in postcommunis countries and the consequences of those designs for particular group in society such as labor, managers, and regional elites raise several im portant questions: Why did some governments and not others includ enormous privileges for enterprise employees in mass privatizatior Why did certain groups in society benefit more than others from th particular form of the mass distribution program? More broadly, wh factors determine the design of privatization programs in postcommu nist states and, by extension, the development of new property regime

In response to such questions, prevailing empirical studies of pos communist privatization tend to emphasize the relative power of poli ical and economic groups in society. Notable examples of this approa are found in the edited volumes of Aslund and in the numerous publ cations of Boycko, Shleifer, and Vishny and of Nelson and Kuze Specifically, the distribution of power in society is said to account f the ability of various groups and stakeholders to acquire special owne ship benefits within a privatization program. Analyzing the speci benefits for groups such as managers, labor, entrepreneurs, and so o these studies contend that the political resistance of economic grou and actors is responsible for the evolution of privatization policies.

Explanations from the privatization literature are consistent with t methodological foundations of broader theories of property rights d velopment from economic history. This body of theory, which is r

[4] Maxim Boycko, Andrei Shleifer, and Robert Vishny, *Privatizing Russia* (Cambridge: MIT P 1995); Lynn Nelson and Irina Kuzes, "Evaluating the Russian Voucher Privatization Program," *C parative Economic Studies* 36 (Spring 1994); idem, *Property to the People: The Struggle for Radical Re in Russia* (London: M. E. Sharpe, 1994); Anders Aslund, *How Russia Became a Market Econ* (Washington, D.C.: Brookings Institution, 1995); Joseph Blasi, M. Kroumova, and D. Kruse, *Kre*

specific to the postcommunist context, characterizes the transformation of ownership systems as a process driven by the efforts of the state to advance particular economic interests (for example, Anderson and Hill, North) or to construct a certain hierarchy of power (for example, Riker and Sened, Riker and Weimer) that is constrained, however, by the distribution of interests and power in society (for example, Libecap, Umbeck, Eggertsson).[5] Thus, an explanation consistent with standard theories of property rights would argue that the inclusion of extensive employee ownership privileges in Russian mass privatization and their absence in Czech mass privatization can be explained by analyzing the power and interests of managers and labor vis-à-vis the state in each country case.

According to this logic, we should have expected equivalent privileges and benefits for management (the industrial *nomenklatura*) and labor in Russia and in the Czech Republic. That is, we should have expected managers and labor to be in equivalent lobbying positions in Russia and the Czech Republic, given that both states emerged from relatively similar property rights structures and industrial structures, with managers and labor having, in principle, equivalent material interests and prior claims to property. Yet the patterns of resistance to privatization and the inclusion of benefits vary enormously from case to case. Without bringing ideology and legitimacy into our analysis, we would realize the differing bargaining positions of these groups vis-à-vis the government only post hoc, namely, with the benefit of knowing how property had in fact been distributed. In such a case the distribution of property would act as both an independent and a dependent variable: it would identify the relative power and interests of groups in society and it would be a product of that power or interest, resulting thereby in tautology.

A crucial shortcoming of standard materialist and power-based analyses becomes especially apparent in comparative analysis; that is, it makes *no attempt to capture the forces (1) shaping how preferences form and (2) determining how the distribution of power is perceived.* At best the preferences and power of various groups are assumed from the outcomes.

[5] Harold Demsetz, "Toward a Theory of Property Rights," *American Economic Review* 57 (May 1967); Terry Anderson and P. Hill, "The Evolution of Property Rights: A Study of the American West," *Journal of Law and Economics* 18, no. 1 (1975); Douglass North, *Structure and Change in Economic History* (New York: W. W. Norton, 1981); William Riker and Itai Sened, "A Political Theory of the Origins of Property Rights: Airport Slots," *American Journal of Political Science* 35 (November 1991); William Riker and David Weimer, "The Economic and Political Liberalization of Socialism: The Fundamental Problem of Property Rights," *Social Philosophy and Policy* 10 (Summer 1993); Gary Libecap, *Contracting for Property Rights* (Cambridge: Cambridge University Press, 1989); Thrainn Eggertsson. *Economic Behavior and Institutions* (Cambridge: Cambridge University Press, 1990).

The inattention to the forces that shape rational expectations and i dividual preference formation can be said to divide the subliteratures o new institutionalism[6] and indeed has been a point of contention be tween proponents and opponents of rational choice theory mor broadly.[7] The critical problem associated with adopting a materialist c power-based approach to studying change in a property regime is tha the standard causal variables—material interests and power—are typi cally treated as black boxes or as simple givens. That is, little attentio is paid, first, to the formation of preferences that shape the content c political lobbying and the demands of groups and, second, to the rea sons behind the relative power of those groups, and by extension, th potency of their political pressure.

Why are simple assumptions about preference formation especiall troubling in the study of postcommunist privatization? It must be mad clear that the assumption of self-interested behavior among actors not directly at issue here; at issue, rather, is the inattention to the def nition of those interests. In a revolutionary environment the formatio of interests and preferences—even those that are purely materialist i nature—is extremely complex, owing to the rapid and dramatic tran formation of political and economic foundational structures. Durir such a period of great uncertainty and flux, identifying which polici best serve one's self-interest or even the greater economic good is f from straightforward for political leaders, industrial specialists, ar economic advisers—not to mention workers, managers, or anyone el whose welfare is affected by the process. This renders problematic methodological approach that exogenizes preference formation whe attributing policy outcomes to the resistance and pressure of stakehol ers, especially stakeholders who simplistically pursue their self-evide objective interest.[8]

[6] The rational choice institutionalist literature, which can include much of the literature on prope rights, offers what has been described as a "thin" or "simplistic" analysis of human rationality and pr erence formation. See the discussion of three subfields within the neoinstitutionalist literature in Pe Hall and Rosemary Taylor, "Political Science and the Three New Institutionalisms," *Political Stu* 44 (December 1996).

[7] For critiques of the rational choice paradigm, see Donald Green and Ian Shapiro, *Pathologie Rational Choice Theory: A Critique of Applications in Political Science* (New Haven: Yale University Pr 1994); Stanley Kelley, "Rational Choice: Its Promises and Limitations," *Critical Review* 9 (Wir 1995); Sven Steinmo, Kathleen Thelen, and Frank Longstreth, eds., *Structuring Politics: Historical stitutionalism in Comparative Analysis* (New York: Cambridge University Press, 1992); Karen Cook Margaret Levi, eds., *The Limits of Rationality* (Chicago: University of Chicago Press, 1990); and J Mansbridge, ed., *Beyond Self-Interest* (Chicago: University of Chicago Press, 1990).

[8] See Oliver Blanchard, ed., *Post-Communist Reform: Pain and Progress* (Cambridge: MIT Pr 1993), chap. 3; and Boycko, Shleifer, and Vishny (fn. 4).

Posing similar difficulties, power-based theories of privatization and property rights development tend to ignore the extent to which the positions of actors in the existing hierarchy of power depend upon shifting human and material resources, as well as upon highly malleable, subjective perceptions of legitimacy and authority. Power is constantly under negotiation, especially during systemic transformation. Given that perceptions of power directly affect the kinds of demands put forth in privatization and the government's willingness to accommodate those demands, the forces shaping the perception and redistribution of power deserve close consideration.

In brief, identifying one's self-interest and the way to best serve that interest is highly complex and uncertain for most if not all economic actors in postrevolutionary environments.[9] Power is shifting rapidly and dramatically, leaving economic and political groups unable to gauge their relative power and perhaps even unable to determine and serve their maximum self-interest. For this reason, we cannot rely on a theoretical approach that explains regime change by beginning with an assumed distribution of interests and power. Rather, we must take one step back and consider how economic interests or preferences form and how the distribution of power is perceived when a property regime is in transition. The intention of this article is to show how ideology determines the formation of preferences and hierarchies of power, which in turn shape the development of postcommunist ownership regimes.

IDEOLOGY AND POLICY DESIGN

In short, to understand why privatization policies extend explicit privileges to certain groups and not to others, it is necessary to adopt a theoretical framework that takes ideology into account. The remainder of this article draws a direct comparison between the Czech and Russian experiences of mass privatization.[10] Despite the obvious differences in

[9] For a discussion of individual strategies for coping with uncertainty in transition, see Michael Burawoy and Kathryn Verdery, eds., *Uncertain Transition: Ethnographies of Change in the Post-Socialist World* (Lanham, Md.: Rowman and Littlefield, 1999), esp. chap. 1.

[10] While this study directs its focus primarily on the mass privatization programs, similar analysis can be conducted for other areas of property reform, such as the privatization of land and housing. In both of these areas the ideological orientations of elite and mass groups have strongly influenced the transformation of property relations. In Russia, for instance, the struggle by the center (that is, Yeltsin's officials) to privatize agricultural land has been both driven and blocked by ideologically opposed actors. Even more so than in the area of industrial privatization, the ideologically charged sparring match between the president's administration and the legislature over privatizing real property has rankled throughout the years of transition. More recently Yeltsin vetoed the Duma's Land Code prohibiting

the scope and complexity of the Russian and Czech economies, i political similarities and equivalent policy starting points make tl well suited for comparison. In both states a small group of liberal a demic economists, having achieved prominent positions in govei ment, developed similar transformative mechanisms to break with t past communist system. Moreover, Russian reformers, who drafti their own reform program shortly after the Czechs, acknowledged th: they had benefited from the prior Czech experience in property reforn Both Russia and the Czech Republic also began their reforms in simi lar international political contexts, with relatively equivalent propert) structures and degrees of etatization.[11] As well, the *proposed* Russiar and Czech programs themselves initially resembled each other in important ways, especially in comparison with the programs of other privatizing states, including other postsocialist states.[12] Both Russia and the Czech Republic implemented a mass privatization program that involved a voucher mechanism, and both distributed property to the population nearly for free. In contrast to the procedure in Hungary, Poland, East Germany, Ukraine, and elsewhere, privatization in Russia and the Czech Republic occurred relatively quickly following the collapse of communism. Furthermore, both teams made immediate privatization a priority. While the similarities between these cases facilitate comparative analysis, it is their difference in the actual (as opposed to the proposed) mass privatization programs that motivates the most interesting questions, specifically, why these two cases differed radically in the extent of privileges offered to economic and regional actors and groups. This article turns to ideology as the variable necessary to answer this immediate question in the hopes of elucidating a much larger area of inquiry: what determines the form of new property rights regimes?

Irrespective of the definition, the term "ideology" necessarily invites controversy. For this reason many scholars avoid the term entirely.[13] For

---

the sale of farmland (July 25, 1997); the Duma then overrode the veto (April 4, 1998). Similarly, the Duma overrode the president's veto of state regulation of agriculture on July 3, 1997. In addition, the federal Duma opposed regional Duma approval of land privatization in various oblasts, such as in Samara Oblast, Tatarstan, and Nizhny Novgorod.

[11] In 1990 in both Russia and the Czech Republic approximately 5 percent of the gross domestic output (GDP) constituted private sector output. In Hungary private sector output as a percentage of GDP was nearly four times greater and in Poland nearly five times greater. For World Bank/OECD figures on etatization, see Liberman (fn. 1), 5.

[12] For explicit discussion of the influence of the Czech mass privatization program on the design of the Russian program, see Boycko, Shleifer, and Vishny (fn. 4), chap. 4.

[13] For instance, Douglass North noted that he prefers the term "belief system" to ideology, since the

ance, several political scientists studying the impact of economic
eoretical constructs such as developmentalism, Keynesianism, and
onomic liberalism on policy-making chose to study the influence of
deas" rather than "ideology."[14]

Although in this article, the term "ideas" is interchanged with "ideologies" for stylistic reasons, the more comprehensive term remains ominant. The term "ideology" can be used in many ways in the social :iences—in fact, in twenty-seven ways, according to one student of leology, Malcolm Hamilton. However, after identifying more than vo dozen of these conceptual elements or "definitional criteria," Iamilton isolates the essential core around which all definitions of idelogy are built: "a system of collectively held normative and reputedly ıctual ideas and beliefs and attitudes advocating a particular pattern of)cial relationships and arrangements, and/or aimed at justifying a parcular pattern of conduct, which its proponents seek to promote, real:e, pursue or maintain."[15]

As the definition notes, ideologies are first and foremost belief sys:ms. Second, ideologies provide an interpretation of the world, both as is perceived (that is, as a Weltanschauung) and *as it should be.*[16] That , ideologies prescribe an ideal and thereby recommend a particular ›rmula for change. Hamilton notes, ideologies "seek to persuade . . . .ıey make claims, present an argument, state reasons, for or against ome plan, programme, behavior, action, conduct, value, attitude, pref'rence, view and so on."[17] Given that the goal of this article is to idenify the mechanisms though which ideologies recommend and shape ›roperty reform, the term "ideology" is preferable to its more neutral 'ounterpart, "ideas."

Moreover, in referring to the term "ideology," the article adopts a Hayekian notion of ideology, using it as a neutral term. Thus, ideology ıere does not imply a Marxian false consciousness but refers simply to . belief system with nothing pejorative intended.[18] Finally, when the ar-

[14] For references see fnn. 51, 52, and 74.

[15] Malcolm Hamilton, "The Elements of the Concept of Ideology," *Political Studies* 35 (March 987), 38.

[16] This aspect (the moral imperative, the ought and not just the is) of ideology is developed in Clif›rd Geertz, "Ideology as a Cultural System," in *The Interpretation of Cultures* (New York: Basic Books, 973).

[17] Hamilton (fn. 15), 36.

[18] When emphasizing the manipulative aspect of ideology some Marxist scholars argue that an idelogy "naturalizes" the political order by masking ideological propositions as truth claims. See Jorge ·arrain, *The Concept of Ideology* (London: Hutchinson, 1979); and the discussion of Larrain and other eo-Marxists in Rhys Williams, "Religion as Political Resource: Culture or Ideology?" *Journal for the*

... to ideology or ideas, it refers not to the idiosyncra[tic] ... of individual leaders but rather to a set of ideas shared by man[y] ... as liberalism, nationalism, anticommunism, or pro-Westernis[m] ... concluding section returns briefly to the choice of focusing on i[deational] ... variables over other nonmaterial variables, such as political c[ulture] ...

THE THEORETICAL FRAMEWORK

This article identifies four primary ways in which ideological f[actors] determine the development of privatization programs and shap[e the] evolution of property rights systems. Empirical examples are [used] throughout to show how ideology is determinative in the follo[wing] four capacities:

1. ideological foundations of theories: ideology determines how privatization programs are drafted
2. ideological context: the ideological context shapes the definition of interests and the distribution of power in society
3. ideological foundations of compliance: ideas influence how leaders attempt to gain support for and compliance to a new property rights system
4. ideological compatibility of programs and peoples: the compatibility between the ideas underpinning privatization policies and ideas and beliefs in society affects the ease of implementation and the distortion of privatization programs over time

These four relationships are expressed in Figure 1.

ELABORATING THE THEORETICAL FRAMEWORK

IDEOLOGICAL FOUNDATIONS OF THEORIES

Ideology determines how programs are designed. Put most simpl[y] there is little doubt that ideology and, in the case at hand, economi[c] liberalism shaped the policymakers' decision to base the new propert[y] system on private ownership. In this regard, ideas as the components o[f] liberal economic theories—which assert that certain economic benefit[s] follow from private property; for example, private ownership leads to more efficient use of resources than state ownership—influence the nature of regime change.

In many respects it is intuitive that economic ideas embedded i[n] economic theory influence economic policy-making. Indeed this occur[s] in two ways. First, such ideas identify an approach in a broad gener[al] ...
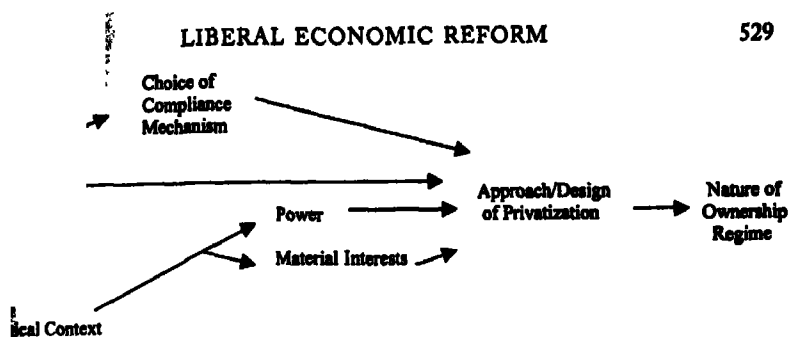
FIGURE 1
THEORETICAL FRAMEWORK: THE CAUSAL MECHANISMS

ense. In this way the anticipated benefits from a liberal property re- orm and the perceived negative consequences of continued state con- rol can recommend that the state find mechanisms to advance the ransformative process as quickly as possible—either to ensure that the goal of systemic change is in fact realized and made irreversible or to eap the benefits of private ownership sooner. In both Russia and the Czech Republic the urgency to privatize was largely motivated by the elief that economic recovery and revitalization could be achieved only y changing the ownership regime.[19]

Second, economic ideas can shape policy in a specific, technical ense; that is, formal economic theories by virtue of their acceptance mong technical experts shape policy-making in specialized arenas. This is a well-known argument previously advanced in the political sci-

[19] The urgency to privatize in Russia stemmed from the fear that the goal of creating a capitalist conomy could be thwarted if it were not accomplished quickly. Given that Yeltsin's reformers under- ood private ownership as the only acceptable institutional arrangement to correct for past structural conomic weaknesses, the achievement of this goal and its irreversibility were paramount. In the Czech epublic there was less fear (although still some) of a return to communism or of the inability to break ee from the Soviet Union entirely. However, the Czech reform team was similarly concerned that the rivatization process must be achieved quickly before various groups, especially managers, "woke up" as ne leading privatizer explains, "and attempted to block the process"; author interview with Dušan říska, Prague, February 13, 1996. The urgency of speed is reflected in the writing of Tříska and Klaus, ho explain their voucher approach in the following way: "Speed . . . was regarded as absolutely es- ential and, therefore, no strategy was regarded as feasible, unless it was capable of producing fast re- ults"; see Václav Klaus and Dušan Tříska, "Review Article of Janos Kornai's "The Socialist System: he Political Economy of Communism," *Dismantling Socialism, An Interim Report: A Compendium of exts from the Years 1992–1994* (Prague, 1994). The article was also published in the Hungarian jour- al *Buksz* (Winter 1994.)

ence literature on epistemic communities.[20] However, those sympa thetic to the role of ideas may take this assertion one step further. On less intuitive and perhaps more controversial level, ideology—in term of normative conceptions of justice—can also be determinative for po icy design. The arguments about the direct influence of widely hel conceptions of justice on policy-making and regime change have bee central to historical discussions of the abolition of slavery in the Unite States and of decolonization in Africa and Asia.

While the extent to which common or elite conceptions of justic shape policy content remains a deeply complex issue, the logic of th argument is straightforward.[21] Therefore, it is useful to turn to the mo subtle ways in which ideology can determine the evolution and desig of property reform.

IDEOLOGICAL CONTEXT

The ideological context shapes the definition of interests and the struc ture of power. There are structural mechanisms through which ideas in fluence the development of property rights. Ideas, in terms of th ideological context, can affect the legitimacy and hence the strength c potential opponents of government programs. As will become clear i the empirical discussion, the "ideological context" is not a fuzzy con cept. It is simply a linguistic shortcut for the sum of the ideas held b members of the elite and mass groups that find expression in both th political discourse and formal institutional mechanisms. In the trans formation of property rights in postcommunist states the ideologica context directly shaped how members of certain groups defined the material interests. Further, it determined the power of key groups t pressure the government to accommodate those interests.

The Czech ideological context following the Velvet Revolution, i which anticommunism was pervasive, weakened the power of certai groups to lobby the government to revise the legislation on privatiza tion. Especially affected in this regard were organized labor and the in dustrial *nomenklatura*, that is, the directors and midlevel managers c state enterprises. Specifically, anticommunism in both institutionalize

[20] On the influence of ideas on policy through epistemic communities, see Peter Haas, ed., *Interna tional Organization: Knowledge, Power and International Policy Coordination* (special issue) *Internation Organization* 46 (Winter 1992).

[21] Given the complexity of this issue and the limited space available here, it is possible to raise on this theoretical assertion. For a detailed discussion of this issue, see Hilary Appel, "Justice and the Refo

and spontaneous forms aggravated the professional insecurity of the political *nomenklatura* and managerial class, leaving them reluctant to obstruct liberal property reform or negotiate changes to the privatization legislation. As a result either of the Czech lustration law[22] or of the spontaneous opposition of midlevel managers and the worker collectives to past political industrial appointments, large numbers of top managers were dismissed from enterprises. Those managers remaining felt that their position was too precarious to challenge the new regime. As Dušan Tříska, the coauthor of voucher privatization and deputy of former premier Václav Klaus, explains, "Managers were taken by every possible group as potential criminals, and everyone took them as representatives of the old regime." Tříska adds that it was consequently not necessary to cater to their interests and let them profit from privatization. Klaus led managers to believe the process was unavoidable.[23] As a result, midlevel managers who survived the forced exodus from industry placed a higher priority on protecting their positions than on trying to reap greater benefits from privatization.[24]

Similarly, the pervasiveness of anticommunist sentiment contributed substantially to the weakness of labor in lobbying the government to assign privileges to enterprise employees. When the privatization legislation was under debate in the parliament, Czech labor organizations were unable to forge the necessary alliances to mount any significant opposition to the privatization programs. At first, the main trade unions—the ČSKOS (the Czech and Slovak Confederation of Trade Unions)—tried to influence economic reform and submitted their own legislation to the parliament. However, following the Velvet Revolution many parliamentarians came to view trade unions as a remnant of the political past and refused to support labor's proposed legislation. In essence, they considered any appearance of support for labor groups in that environment to be politically risky.[25]

[22] In the Czech Republic the lustration law was an anticommunist screening measure that required the bureaucratic and industrial elite to resign from certain top posts for past acts of political collaboration, in particular with the communist secret police. Vojtech Cepl, "Lustration in the CSFR," *East European Constitutional Review* (Spring 1992); Jirina Siklova, "Lustration or the Czech Way of Screening," *East European Constitutional Review* (Winter 1996).

[23] Author interview with Tříska (fn. 19).

[24] See Hilary Appel and John Gould, "Identity Politics and Economic Reform: Examining Industry-State Relations in the Czech and Slovak Republics," *Europe-Asia Studies* 52 (January 2000).

[25] For proclamations by trade union leaders that parliamentary deputies were intentionally taking antiunion stands, see *ČTK National News Wire*, December 6, 1990. On the difficulties of finding support from the center of Parliament (including from Civic Forum) on labor issues, see *ČTK National News Wire*, December 4, 1990. For further details of the trade unions' position on privatization, see *Práce*, November 6, 1990. Difficulties on the part of labor in finding support in parliament were also

Regarding mass privatization in particular, labor unions were unabl to convince the government or the parliament to reserve a significan block of shares for employees within the program of voucher privatiza tion, even though the first legal document proposing mass privatiza tion, the Scenario for Economic Reform,[26] referred to employe ownership as a standard method that could be implemented alongsid voucher privatization. The very weakness of labor is underscored by th workers' inability to acquire rights to even the legal amount allotted t them. In fact, the percentage of shares reserved for employees turne out be much lower in practice than the 10 percent (later amended to percent) maximum allowed by law.[27]

Despite the attempts by the new trade union leadership to she labor's reputation as a communist holdover, the formal restructuring o the unions[28] went unrecognized during this crucial early period i which the privatization program was adopted. Only those members o parliament who were clearly on the left (for instance, from the Com munist Party or *Obroda*) were interested in supporting labor. Ironicall their support was entirely unwelcome, and ČSKOS avoided such affilia tions as part of its ongoing efforts to break with its communist past. I other words, for the same reasons that Czech parliamentarians pre ferred to avoid any association with labor so as not to appear too far lef on the ideological spectrum, labor was also hesitant to form alliance with communist deputies. Hence, the anticommunist beliefs of labo leaders and parliamentarians prevented coalitions from forming to ef fect changes within the privatization program.

Post-Soviet Russia did not experience formalized manifestations o anticommunism like those in Czechoslovakia following the Velvet Rev olution. The continued legitimacy of the legislature elected under th communist regime and the absence of anticommunist screening law (that is, lustration laws) ensured the professional security of past powe

trepeated in author interviews with members of the first postcommunist Federal Assembly, includin Jan Kavan, Prague, February 7, 1996, and with members of the Czech Parliament, including Václa Řak, Prague, February 22, 1996.

[26] See provisions on preferential sale to employees in excerpts from the Scenario for Economic Re form, in "Document: Ze Scénáře Ekonomické Reformy," *Lidové noviny*, September 4, 1990, 2.

[27] See Josef Kotrba, "Privatization Process in the Czech Republic: Players and Winners," in *Th Czech Republic and Economic Transition in Eastern Europe* (San Diego, Calif.: Academic Press, 1995) 164–65.

[28] Over 80 percent of former trade union leaders had been dismissed, 90 percent of the member had given up their Communist Party membership immediately after the Velvet Revolution, and, more over, the ČSKOS prohibited all formal affiliations with political parties in enterprises and workshops See "Reinventing Trade Unions" (Manuscript, Central European University Privatization Project, Fal

holders. The implementation of Russian privatization was confounded by the absence of formal mechanisms that could mute the parochial demands of the former *nomenklatura* or complicate the formation of antigovernment or antiliberal alliances. In other words, because parliamentarians elected under the communist regime or managers appointed under the previous system were less troubled by professional insecurity, they were able to make bolder claims for property. As a result, the privatization team confronted enormous political pressures and in response sought political support in exchange for granting widespread exemptions and special privileges to sectoral and factoral groups. As the privatization process advanced, the ability of Russian privatization officials to build support for their economic agenda became even more dependent on enormous transfers to potential supporters within the industrial and financial elite, peaking with the 1995 loans for shares program.

In essence, one must recall that when the 1992 Russian privatization program was being debated, Russian privatization officials were working within a different ideological context and thus confronted a legislative arena that contrasted sharply with that faced by Klaus and the Czech property reformers. Unlike in most other postcommunist states in Eastern Europe, there were no new elections called in post-Soviet Russia; and those politicians elected under the previous system were deemed legitimate within the new polity. The lack of parliamentary support is evident as early as January 1992, when the speaker of the parliament called on the government to resign.[29] Similarly, those managers in top posts in Russia (many of whom had obtained their positions for political reasons in the past) maintained their positions. Hence, when the privatization debates began in the legislature, the Supreme Soviet included a high proportion of communist deputies[30] who were ideologically sympathetic to employee rights, or at least politically compelled (to appear) to support them.[31] As a result, Russian managers were able to form alliances with members of the federal legislature and pressure the State Property Committee to increase their ownership stakes in privatization under the banner of employee rights.

[29] *Izvestiia*, January 14, 1992, 1.

[30] On the ideological makeup of the highest political organs from a progovernment perspective, see Al'bert Plutnik, "Atakuia pravitel'stvo, oppozitsiia boitsia uspekha reform," *Izvestiia*, April 9, 1992, 1–2. For estimates of the percentage (87 percent) of members of the Supreme Soviet believed in 1992 to be former communists, see *Finansovaia Izvestiia*, November 18, 1992, i.

Initially, Anatoly Chubais, as the chairman of the committee, r
sisted the transfer of property to both managers and rank-and fil
workers.[32] This, however, did not prevent Russian employees fro
obtaining from the 1992 Russian Privatization Law the controllir
stake of almost three-quarters of enterprises privatized.[33] Leading pr
vatization officials openly and not infrequently lamented the extent
employee ownership and described it as a serious flaw in the progran
design.[34]

Despite the compromise, Russian privatization officials at the san
time directed their efforts toward indirectly limiting the amount
ownership control given to workers.[35] For example, Chubais insist
that worker shares be fully transferable and individually owned rath
than collectively owned by the trade union or the enterprise worke
council. Full transferability was expected to facilitate the movement
worker shares.[36] Owing to the transferability of worker shares ar
vouchers and the maneuvering of managerial and industrial elite
worker ownership declined somewhat over time while managerial co
trol over enterprises soared.

To summarize, in these instances, contrary to a power-based or m
terialist interpretation of politics, ideas have influenced structures
power and the definition of material interests rather than the revers
In the Czech ideological context anticommunism served to discred
labor demands and made managers reticent, thus preventing these tw
groups from shaping the design of privatization. Moreover, the prev

[32] The government first proposed that employees receive 25 percent of the shares of their enterpr for free; however, these shares would be nonvoting. Yet as greater pressure mounted, the reform te offered a second and then a third variant from which worker collectives could choose. In the seco variant employees could buy 51 percent of the voting shares for a negligible price. In the third varia managers or a small group of workers could buy 20 percent of voting shares at the nominal price, b several conditions governed the transaction. The second variant was designed to placate workers a managers and the third variant was specifically intended to satisfy managerial demands. The third va ant was seldom applicable, though, due to strict eligibility requirements that were obscurely writt into the legislation. In addition to these three variants, a fourth one was proposed by the Parliame but successfully thwarted by the reformers—to transfer up to 90 percent of the shares of an enterpr to the worker collective. For further detail on the three variants, see Roman Frydman et al., *The P vatization Process in Russia, Ukraine and the Baltic States* (London: Central European University Pre 1993), 53–58; and on the fourth variant, see Darrell Slider, "Privatization in Russia's Regions," *Post-S viet Affairs* 10 (October–December 1994), 375.

[33] See fn. 2.

[34] Dmitry Vasiliev's assessment of employee privileges, reported in *Financial Times*, November 2 1993, 2. See also former Russian State Property Committee chairman Alfred Kokh's depiction Kokh (fn. 2), 51, 82.

[35] For an extended discussion the Russian government's intentions and the distributive cons quences, see Appel (fn. 3), 1433–49.

[36] See Andrei Shleifer and Dmitry Vasiliev, "Management Ownership and Russian Privatizatio

lence of anticommunism influenced the preference formation of Czech managers who, in contrast to their Russian counterparts, preferred to minimize the risk of losing their job over becoming proprietors of their firms. In Russia anticommunism was less prevalent in the political discourse and did not find anywhere near the same level of expression in formal institutions. As a result, the legitimacy and hence the power of many groups who had benefited from the past communist regime was strengthened (or at least was not weakened), which in turn affected their ability to advance their claims to property during the reformulation of Russia's ownership regime.

IDEOLOGICAL FOUNDATIONS OF COMPLIANCE

Ideas influence how leaders attempt to gain support for and compliance with the new property rights system. For the sake of implementing privatization programs, the ideas of leaders affect the way they approach the task of eliciting support for and assuring compliance with newly designed policies. In investigating how leaders go about eliciting compliance from groups in society, it is useful to borrow Amitai Etzioni's theoretical framework from organizational theory.

According to Etzioni, the state (or any organization) can employ three types of mechanisms to elicit compliance from target populations: coercive, remunerative, and normative.[37] Coercive mechanisms involve a credible threat or the actual use of force to gain compliance, such as the threat of violence or incarceration. Coercive mechanisms are not particularly helpful for understanding how postcommunist governments achieved compliance with the new private property regime. However, governments have relied throughout history on coercive mechanisms to found new ownership regimes, a key case in point being the founding of the communist system of state ownership. During the 1930s coercion enabled the nationalization of property, collectivization, and dekulakization in the Soviet Union.[38]

Another set of mechanisms employed to elicit compliance with a private property regime is remunerative or utilitarian in nature. Such mechanisms rely upon economic benefits and special privileges to win

[37] Etzioni, *A Comparative Analysis of Complex Organizations: On Power, Involvement and Their Correlates* (New York: Free Press, 1961), 34–22. See the discussion (and extension) of the three reinforcing mechanisms in Ian Lustick, "Hegemony and the Riddle of Nationalism: The Dialectics of Political Identity in the Middle East," Working Paper 1997-01 (Christopher H. Browne Center for International Politics, University of Pennsylvania, 1997).

[38] Jerry Hough and Merle Fainsod, *How the Soviet Union Is Governed* (Cambridge: Harvard Uni-

support, such as special enterprise share-holding privileges, bribes monetary transfers.[39] *The free transfer of the shares of formerly st owned enterprises to the population through vouchers is an exampl* a utilitarian compliance mechanism because it helped to ensure the p *ticipation and support from the population at large.*

The third type of mechanism, based on ideological reinforcement also important for gaining support for a new ownership regime. In t case, affected groups in society support a program not because th directly benefit, but because they consider it legitimate, fair, appropriate.

The value of borrowing from Etzioni goes beyond simply being al to assert that postcommunist governments rely upon ideological leg imization. This is intuitive as well as consistent with mainstream the ries of change in property rights and policy-making more broadly Moreover, few would disagree that governments make ideological a sertions, even if only to justify or mask a particular agenda. *Rather, importance of compliance mechanisms for understanding the role of ideas emphasized in order to argue that leaders' ideological beliefs determine wh among the three reinforcing mechanisms they employ.* This choice does n merely represent an intellectual exercise by political strategists. Instea this choice can affect the strength of political resistance as well as t content of a policy program—especially when leaders chose utilitari mechanisms requiring material benefits to be incorporated into program.

While both the Russian and Czech leaders employed remunerati reinforcing mechanisms such as the free distribution of proper through vouchers, the Czech leaders in addition relied extensively anticommunist and pro-European ideas to develop highly effective id ological reinforcing mechanisms. Throughout the privatization pro gram, Klaus recognized the need to maintain his support base throug an ongoing political effort to sell and legitimate the transformatic process.[41] His own comments acknowledge this deliberate campaign

[39] For a discussion of the dependence of normative compliance mechanisms on material incenti and coercion potential, see Margaret Levi, *Of Rule and Revenue* (Berkeley: University of Califor Press, 1988), 68.

[40] For instance, North considers ideology primarily as an instrument of the state to justify progra and as a tool of the opposition when attacking existing arrangements. Ideas do not drive the formati of property rights systems, but they have played an important role in the maintenance of and challen to existing property rights. North (fn. 5), 50–52.

[41] Quoted from the text of the speech by Klaus at G-30 conference in Vienna on April 24, 19 "The Czech Republic's Prospects . . .," *Telegraf,* May 4, 1993, cited in "Klaus Hands Down 'T Commandments' for Reform," *FBIS-EEU* 93–087. May 7. 1993. 7. 9

ing style and show the extent to which he recognized the value of reinforcing mechanisms. Klaus states:

> It is absolutely essential for reformers to have faith in the success of the reforms to be able to inspire all their fellow citizens and be able to establish broad pro-reform coalitions. The reform is not an academic problem: it is a political affair and it is of immense importance to enlist sufficient political support for it. We realized at the very beginning that establishing a political platform was an indispensable part of reform activities. . . . This required immense political and human efforts, hundreds of political tours, and endless meetings with thousands of people—in other words, it required, and still does, a permanent campaign.[42]

Attempting to legitimate his economic approach with ideologically charged rhetoric, Klaus presented his program to the greater population as the most liberal approach—one that would restructure the former Soviet-style planned economy according to "European" or "Western" economic principles.[43] More specifically, Klaus was very effective in selling the historical appropriateness of the approach. His program would "return" the country to its rightful place as a member of the Western or European communnity. In this way he equated his pro-European emphasis with a pro-Czech position.

In Russia, by contrast, rather than developing ideological legitimating mechanisms in the tradition of Klaus, reformers did little to promote the mass privatization process on ideological grounds. Officials relied instead upon material incentives to individuals and groups to implement privatization and eschewed ideological campaigns and political slogans in their writings and public statements. As Chubais, the architect of Russia's mass privatization program, explained the future viability of the program: "As it has been planned . . . each citizen of Russia will in reality become a proprietor, not at the level of slogans and appeals, but at the level of a monetary document in his pocket. The implementation of this program will itself become an irrevocable guarantee for [a] more comprehensive process of reforms."[44]

As Russian privatization officials themselves acknowledge, they sought to advance the process, first, by buying popular support with a voucher mechanism and employee benefits and, second, by co-opting powerful groups and sectors by granting numerous concessions, privi-

[42] Ibid.

[43] Zdenka Mansfeldová, "Professional and Political Strategies in Economic Discourses" (Paper presented at the conference on Symbolic Politics and the Process of Democratization in Eastern Europe, Berlin, August 21–28, 1994).

[44] "Press Conference with Anatoly Chubais," *Official Kremlin International News Broadcast*, July 30, 1992.

leges, and exemptions. In a speech to the Congress of People's Dep ties, Chubais stated:

> A legal framework has been laid down to get privatization really moving. It is essential to this end to achieve a correct balance of interests of all social groups involved. . . . The privileges and exemptions provided under our privatization program are the largest in the entire history of privatization throughout the world. At the same time we are convinced that it is equally impossible to get privatization moving without providing effective incentives.[45]

Why did the Russian reformers rely primarily on material reinfor ing mechanisms to sell the privatization program and stay away fro using Russian cultural symbols or historical referents? To answer th question, one must consider the ideas of these actors, in particular, tl liberal economic ideas and professional identities of the Russian arch tects of privatization. Yeltsin's property reformers, that is, were con mitted to the liberal economic principle that all economic beings a rational self-interested actors who respond intuitively to material ii centives. This liberal (or neoclassical) ideological disposition is exen plified by the scholarly work of Maxim Boycko, a leading economist c the privatization team. He wrote explicitly that there was nothir unique to the Russian nation that would to preclude the implement: tion of a liberal economic regime.[46] According to this logic, any can paign linking privatization to specifically Russian cultural or historic symbols or any revision of the program to make it appear particular suited to the Russian context would imply that Russians do not re spond rationally to material incentives.[47] Furthermore, any suggestio that the Russian nation somehow fell outside the liberal economic pa adigm would undermine the universal premise of their liberal ideolog

To be clear, the ideological vision of Yeltsin's reformers was not sub stantially different from Klaus's vision of integrating the economy an

[45] "Speech of Anatoly Chubais to Congress of People's Deputies," *Official Kremlin Internation News Broadcast*, April 8, 1992.

[46] Boycko was a key player in Russian property reforms, holding numerous positions in the privat zation process, including director of the Russian Privatization Center and briefly chairman of the Sta Property Committee (also known as the minister of privatization). On the universality of the ration actor model, see Robert Shiller, Maxim Boycko, and Vladimir Korobov, "Hunting for Homo Soviet cus: Situational versus Attitudinal Factors in Economic Behavior," *Brookings Papers on Economic Acti ity* (Spring 1993); and idem, "Popular Attitudes toward Free Markets: The Soviet Union and tl United States Compared," *American Economic Review* 81 (June 1991).

[47] As Piotr Aven, an economist working with Gaidar and minister of economic relations, writes c this theme: "There are no special countries. All countries from the point of view of an economist a the same in what concerns the stabilization of their economies"; in *Nezvisimaia gazeta*, March 2 1992, quoted by Vladimir Shlapentokh, "Privatization Debates in Russia, 1989–1992," *Comparati Economic Studies* 35 (Summer 1993), 19–32

the country with the West. The insistence by Russian privatization officials that Russia—just like the capitalist world—was populated by self-interested economic agents underscores their Westernizing project. However, it was not on this basis that they attempted to build public support within the domestic political arena.

In the early period of privatization Yeltsin's reformers simply did not see it as their place to launch a cultural campaign, since they thought of themselves as economic technocrats and not politicians—a distinction they themselves drew.[48] Undertaking an explicitly ideological campaign to build popular support was distasteful to the reformers, as such a campaign would have smacked of communist-era tactics. Yeltsin, in his memoir, criticizes the liberal economic reformers for not wanting to "dirty their hands with politics."[49]

Thus, although both the Czechs and the Russians sought to make their programs appear fair to the greater population, only the Czech reformers attempted to link the creation of the new property regime to the founding of a postcommunist national identity. The Russian reformers for their part eschewed directing any ideological reinforcing mechanisms at the domestic audience. At best Chubais (like Klaus) appreciated the value of appearing pro-Western and liberal for *international* audiences. In the end, failing to focus on ideological legitimation in Russia greatly hindered the liberals' attempts to implement and sustain the privatization program and ultimately led them to concede certain privileges to factoral, sectoral, and regional groups against their preferences and even against some of the liberal ideals that inspired the construction of a new property regime.

IDEOLOGICAL COMPATIBILITY

The fourth way in which ideas influence the implementation builds upon the first three. Namely, a lack of compatibility between the ideological basis of a program and the ideas of elite and mass groups increases the cost of political reinforcement; and depending on the will and skill of leaders, this incompatibility can determine the extent to which liberal reform is distorted.

The notion of an ideological compatibility between a program and a particular context has its roots in Weber's *The Protestant Ethic and the*

[48] On the reformers' self-perception as macroeconomic technocrats indifferent to popularity ratings, see "Egor Gaidar i v bezbykhodnykh situatsiiakh nado iskat' vykhod," *Izvestiia*, July 5, 1992, 1, 3.

Spirit of Capitalism.[50] More recently it has been studied by such pror nent political economists as Peter Hall and Kathryn Sikkink, who fo on the role of ideas in the introduction of new economic regimes.[51] A though these scholars do not speak of reinforcing mechanisms per they do note the importance of what they call the "fit" between ideas specific groups in society and the ideas underpinning new econon paradigms for how those paradigms are received and supported.[52]

In this regard, the importance of compatibility relates to the politi resistance that reformers face and the cost of social compliance and p litical reinforcement. *In particular, the less the ideas underpinning a p gram fit effortlessly with the ideas of different economic and political grou the greater the cost of developing reinforcing mechanisms.*

A comparative study of how postcommunist governments prom liberal property reform illustrates this principle well. In the Czech R public the reformers strengthened the case for privatization by portra ing it as anticommunist (for example, with the Property Restituti laws) and pro-European (with the institutionalization of private syst of ownership)—and thus they argued, essentially Czech. It was rel tively easy (that is, it took little effort) for the Czech reformers to arg this logic by emphasizing that the communist system had been ir posed and enforced from without and needed to be replaced by a sy tem more appropriate to the Czech context, namely, a Europe system. Given existing territorial and cultural referents, it was rath straightforward for Klaus to make the creation of a new property sy tem part and parcel of the formation of a new Euro-Czech identity.

This pro-Europe/pro-Czech campaign was not limited to Klau even if he at that time attempted to appropriate it as his own.[53] The f mous slogan "Return to Europe!" was Civic Forum's successful slog in the June 1990 elections. By including the word "return," this slog

[50] Max Weber, "'The Protestant Ethic and the Spirit of Capitalism," in W. G. Runciman, ed., *We Selections in Translation*, trans. Eric Matthews (Cambridge: Cambridge University Press, 1968).

[51] Hall, *The Political Power of Economic Ideas: Keynesianism across Nations* (Princeton: Princeton U versity Press, 1991); Sikkink, *Ideas and Institutions: Developmentalism in Brazil and Argentina* (Itha N.Y.: Cornell University Press, 1991).

[52] The notion of ideas "fitting" or "resonating" in a given context as an important insight from ideas literature can be found within the works of many influential contributors in one form or anoth See Hall (fn. 51); Sikkink (fn. 51); and G. John Ikenberry, "Creating Yesterday's New World Ord Keynesian 'New Thinking' and the Anglo-American Postwar Settlement," in Judith Goldstein a Robert Keohane, eds., *Ideas and Foreign Policy: Beliefs, Institutions and Political Change* (Ithaca, N. Cornell University Press, 1993).

[53] See Klaus's remarks when forming the Interparliamentary Group of the Democratic Right Wi a subgroup of Civic Forum, and declaring the fundamental principles of the organization

einforces the already prominent notion of Czechoslovakia as part of urope and implicitly reaffirms prevailing conceptions of geopolitical livisions. (After all, the slogan could have also been "To Europe!") Inleed pro-Europeanism resonated—or at the very least was made by oliticians to resonate—with current popular notions of history and erceptions of geographical spheres and with cultural and religious diides, all of which were evoked to justify the nation's place in Western ather than Eastern or Slavic civilization. Yet despite the existing pro-uropean sentiment, the fact that Klaus chose to stress a more Western nd more integrative approach, rather than fueling a strictly nationalist lentity was in no way preordained. Consider, by contrast, the effort by lovak premier Vladimír Mečiar to foster a more nationalist identity luring the same time period. Nonetheless, given the Czech tradition of sserting its place in Europe and the popularity of existing slogans, it vas not difficult for Klaus to present the property structures of a Westrn, liberal system as essentially Czech and thereby to help legitimate is liberal economic approach.[54]

At the same time Klaus placed himself in stark opposition to anyhing that smacked of a compromise between Western capitalism and oviet communism. He sought to link the privatization program to a idespread desire to break with the past, which meant dismantling the ommunist system and differentiating the country from the Eastern loc. In the process of defining himself as the leader who would return he Czechs to Western Europe, he also implied that those less committd to the free market would hold the country back, keeping it in the ommunist era. In order to highlight the Western liberal nature of his pproach, Klaus would label the alternative programs as "third ways," reform communist," or simply "socialist." By contrast, he asserted that e "want[ed] a market economy without any adjectives," insinuating hat the society must choose between market capitalism and another ersion of communism.[55] Klaus's view is encapsulated in an address to ivic Forum: "No path of socialism with a human face, no left alternaves or Obroda [communists expelled from the party after the 1968 oviet-led invasion of Czechoslovakia] can lead to a modern European tate. We are not returning to the year 1968, the key year is 1948. . . . W]e are following the tradition of the First Republic (1918–1938),

[54] Similarly, in building a new, postcommunist identity, political leaders resorted to similar appeals history, citing the Masaryk period in particular, to support the "naturalness" of democracy in the zech lands. For further analysis, see Ladislav Holy, *The Little Czech and the Great Czech Nation*

Christianity, European Civilization."[56] A similar strategy would of course have been substantially more complicated in Russia, given its very different ideological context and territorial referents. Since the inception of market reforms, the rejection of the communist past in favor of a new Western liberal orientation has often been interpreted in post-Soviet Russia as a rejection of oneself and demeaning to one's past. And while history is certainly subject to reinterpretation and distortion, it would require a substantial effort to make a persuasive case in Russia that the adoption of capitalism (and private property) was a return to a former self. Russian politicians could not reject communism as alien or overlook the foreign connotation of a liberal property system with anywhere near the same ease as Czech politicians could. Furthermore, promoting such a campaign would have required not only willingness on the part of the liberal reformers but also sufficient charisma and skill.

COMPATIBILITY, LEADERSHIP, AND ONE FINAL CAVEAT

This emphasis on the importance of a compatibility—or in Weberian terms, an "elective affinity"—is not meant to suggest that the ideological beliefs of dominant groups are fixed or that the ideological significance or symbolism of a new economic project is immutable. Through the manipulation or the repackaging of a project, this fit between the ideas underlying a new ownership regime and the ideas in society can be improved. This depends primarily on which aspects of the process are emphasized (or deemphasized) or on which salient changes to the legislation are adopted. Similarly, if one takes a Gramscian perspective, the dominant ideas themselves in the political environment can be rearticulated in a way more in line with liberal economic reform and the privatization process.[57]

[56] *ČTK National News Wire*, December 8, 1990. Examples of Klaus's evocation of history in public speeches abound. Note Klaus's historical references when he founded Civic Democratic Party (ODS) and announced the programmatic principles of the party. According to the Czech national news agency, Klaus explained that ODS would "follow the traditions of European Christian civilization, the humanistic and democratic traditions of the preordained-Munich republic (1918–1938) and the experience of the present Western democracies. It resolutely and entirely rejects Marxist and Leninist ideology, and all trends toward Socialization and Collectivization in the economy and politics are alien to it." Applicants for membership in the ODS would have to state whether they had been members of the Communist Party, and applications from former members of the People's Militia and collaborators with the former secret police would be rejected; *ČTK National News Wire*, March 1, 1991.

[57] For a discussion of the Gramscian concept of "articulation," see Grossberg, who defines articulation as a "continuous struggle to reposition practices within a shifting field of forces, to redefine possibilities by redefining the field of relations—the context—within which the practice is located." Lawrence Grossberg, *We Gotta Get Out of This Place: Popular Conservatism and Modern Culture* (London: Routledge, 1992), 54, quoted in John Kurt Jacobsen, "Much Ado about Ideas: The Cognitive Factor in Economic Policy," *World Politics* 47 (January 1995), 308–9.

The malleability of social norms and mass perceptions raises important questions about the role of ideas. If programs can be packaged to fit a given context, how does the ideological context constrain the development of property rights structures? Similarly, if widely held ideas and beliefs are subject to manipulation, how important is the issue of compatibility? More broadly, what power does ideology ultimately wield? *Ideology will constrain and distort the implementation of a privatization program when the ideas of a program do not effortlessly fit with the ideological context and when political entrepreneurs lack the skill or the will to develop ideological reinforcing mechanisms that avoid changes within the structure of the program.*

As discussed above, leaders' beliefs may prevent them from developing effective reinforcement mechanisms, as occurred in Russia, where privatization officials refused on principle to mount an ideological campaign for mass privatization. Even more commonly, though, new leaders may simply lack the political skill to construct effective ideological reinforcing mechanisms.[58] Either they fail to associate the project with positive symbols, or they cannot find a language to describe a program that avoids alienating mass groups. Alternatively, they may neglect to find those elements in the program that relate to the issues most salient in the political discourse and may inadequately emphasize how the project reinforces cultural or territorial identities. Consequently, when there is no immediate resonance between the ideas behind privatization and the ideas of major groups in society, and when political entrepreneurs cannot construct effective ideological reinforcing mechanisms, then the incompatibility between the ideas of a program and the ideological context has a generative effect on policy content by altering and hindering the realization of the new property regime.

In sum, through the development of an ideological legitimating campaign, the Czech reformers were able to bolster the privatization process by equating the institutionalization of a European or Western property regime with a return to a former self. Klaus played on the notion that a European property system was organic to the Czech soil and that its (re)establishment could be understood as part of the process of rebuilding what had been repressed. On the contrary, in post-Soviet Russian society European or Western symbols and policy programs lacked an equivalent preexisting nationalist connotation (save a few ex-

[58] John Hall writes: "The discovery that intellectuals can construct social identities should not be exaggerated. . . . [I]t is important to remember that the exercise of this type of ideological power is rare. . . . More important, normative cohesion is, given social complexity, almost certain to be incomplete"; Hall, "Ideas and the Social Sciences," in Goldstein and Keohane (fn. 52), 53.

ceptions such as in St. Petersburg), thereby increasing the cost of political reinforcement.

In Russia the loss of an empire and an ideology left the Russian identity in tatters; hence the establishment of undeniably Western structures—without developing an alternative ideological legitimating mechanism—could easily have been portrayed as one more blow to the Russian sense of place in the world. While an ideological campaign in Russia based largely on Western liberalism most likely could not have enabled liberal reformers to avoid transferring a vast amount of shares to employees and the former industrial *nomenklatura*, especially given its high cost in terms of human effort and skill, there is no reason to preclude the possibility of a differently designed, ideological reinforcing mechanism being effective. The liberal ideas that underpinned the reformers' privatization program could have been repackaged to appeal to the dominant culture—still avoiding a Western connotation but appealing instead to the Russian desire to be seen as part of the "civilized" world—a word that reverberates throughout Russian popular discourse.

Concluding Thoughts and Theoretical Applications

Throughout much of this article the prevailing ideological context within a country at a specific historical juncture could potentially be described in cultural or, more narrowly, in *political-culture* terms, if it were not for one crucial aspect. Namely, "political culture" traditionally connotes a deep-rooted persistence of behavioral patterns. Although the literature on political culture is extensive and diverse, political culture is typically understood in relatively static terms.[59] Some even imply that a country's political culture is essentially immutable.[60] Characteristic of this approach, the historian Richard Pipes contends in reference to Russia that political culture changes as slowly as language.[61]

Political cultures have often been understood as resistant to change ever since Gabriel Almond and Sydney Verba popularized the term in

[59] On the culture-ideology distinction as representing the stasis-change dichotomy in the work of Clifford Geertz, Ann Swidler, and others, see Rhys Williams, "Religion as a Political Resource: Culture or Ideology?" *Journal for the Scientific Study of Religion* 35 (December 1996).

[60] For an analysis of political culture as the product of the protracted historical accumulation of practices and meanings, see Thomas Hansen, *The Saffron Wave: Democracy and Hindu Nationalism in Modern India* (Princeton: Princeton University Press, 1999), 27.

[61] Pipes writes: "Political Culture, shaped by a nation's historic experience, enters the nation's bloodstream and changes as slowly and reluctantly as does language or customs"; Richard Pipes, "The Communist System," in Dallin and Lapidus, eds., *The Soviet System in Crisis* (Boulder, Colo.: Westview Press, 1991), 18.

the early 1960s in their study on the determinants of democracy.[62] In discussing the type of political culture essential for democratic development, Almond and Verba emphasize that a "civic culture" in democratic polities takes centuries to develop.[63] Political culture develops through the fusion of attitudinal and behavioral orientations: new patterns cannot replace old patterns, they can merely *fuse with existing patterns*. Thus, while a political culture favorable to democracy can be encouraged through the creation of voluntary organizations, socializing agencies, or the mass education system, the transformation is necessarily a protracted one.

The emphasis on the *persistence* of a political culture has invited substantial criticism. For instance, Philippe Schmitter and Terry Karl emphasize the limitations of relying on a state's political culture to explain political phenomena. In discussing the necessity of the existence of a civic culture (or "ingrained habits of tolerance, moderation, mutual respect . . . or trust in public authorities") for democracy to take hold, Karl and Schmitter write, "Waiting for such habits to sink deep and lasting roots implies a very slow process of regime consolidation . . . and would probably condemn most contemporary experiences *ex hypothesi* to failure."[64] Criticized on many fronts, the concept has been described by David Laitin, a prominent contributor to the debate, as having been in the "professional doghouse" for years.[65]

Although many political scientists have sought to revive the concept of political culture along new lines, the term often maintains the static connotation of historical rootedness and persistence in the face of change in a way that does not hold for ideology-based variables.[66] Analyzing a specific context in ideological rather than cultural terms allows a larger role for leadership and implies much greater dynamism. Thus, while external shocks, systemic crises, or even charismatic leaders could profoundly affect the salience of competing ideologies, they would not

[62] Gabriel Almond and Sidney Verba, *The Civic Culture* (Boston: Little, Brown, 1964).

[63] Ibid., 500.

[64] Schmitter and Karl, "What Democracy Is and Is Not," *Journal of Democracy* 2 (Summer 1991), 83–84.

[65] See Laitin, "Compliance Ideologies: Rethinking Political Culture," *American Political Science Review* 89 (March 1995).

[66] On the revival of the study of political culture in the second half of the 1980s, see Ronald Inglehart, "The Renaissance of Political Culture," *American Political Science Review* 82 (December 1988). More recently, see Richard Wilson, *Compliance Ideologies: Rethinking Political Culture* (New York: Cambridge University Press, 1992), as well as Laitin's analysis of political culture in Laitin (fn. 65), 168. Additionally, see Larry Diamond, *Political Culture and Democracy in Developing Countries* (Boulder, Colo.: Lynne Rienner, 1994), especially the introductory chapter. On Eastern Europe, see Vladimir Tismaneanu, ed., *Political Culture and Civil Society in Russia and the New States of Eurasia* (Armonk, N.Y.: M. E. Sharpe, 1995).

lead to the rapid *abandonment* of a nation's political culture. Hence, the choice to focus on ideology over culture in this article is predicated on the understanding that ideas and ideologies are more readily subscribed to and, by extension, more easily abandoned than political culture.

Studying the ideological context, however, does not entirely evade the problems associated with studying nonmaterial determinants of political change and certainly the term "ideology" carries so much baggage that many scholars prefer to avoid it altogether. To cope with some of these problems, this article uses a narrow definition of ideology, as discussed above. It also considers the ideological context identifiable only when a particular ideology finds formal expression in public debates *and* in tangible institutions. For instance, "anticommunism" as an ideology is manifested not only in popular and political discourse but also in formal lustration (screening) laws and restitution laws; likewise, not only is "pro-Europeanism" as an ideology represented and repeated in public speech and opinion survey data,[67] but it is also institutionalized in the formal project to join the European Union.

One issue that arises, however, in describing the choice of a privatization strategy as ideologically driven rather than culturally driven is whether all relevant actors, or even key actors, must genuinely subscribe to that ideology for it to be determinative. Joseph Schull and others who study ideology in linguistic terms contend that the subscribers to a shared ideology need not (and typically do not) share identical beliefs.[68] Moreover, Schull maintains that the intensity of belief is not the most important measure of the power of an ideology. The determinative power of an ideology rests upon the constraints it places on political debates and the consideration and articulation of alternatives. Thus, in this counterintuitive but compelling argument, Schull contends that the power of an ideology should not be measured by the sincerity of its advocates (since methodologically this is impossible to gauge) but by the commitment to the ideology as revealed consistently in speech and behavior.[69]

[67] Public attitudes in East European countries toward Western Europe can be found in the *Central and Eastern Eurobarometer*, nos. 1–6 (Brussels: European Commission, 1990–96).

[68] On the "unwarranted assumptions about the uniformity of convictions" among members of an ideological group, see Schull, "What Is Ideology? Theoretical Problems and Lessons from Soviet-Type Societies," *Political Studies* 40 (December 1992), 728–31. Hansen (fn. 60) notes the tensions created by the "multiplicity of meanings inscribed in most ideological constructions. This tension may be expressed as the tension between the conceptual grammar of a discourse and the connotative domain within which it is articulated" (p. 25).

[69] Schull notes (fn. 68) that it would be mistaken to "put a premium on the genuineness of an agent's state beliefs" since "one's actions will be shaped by an ideology in so far as one must conform to its conventions"; one need not "believe in the ideology... one must be committed to it. The required attitude is respect, not faith." For this reason, Schull suggests that ideologies are better understood as a discourse rather than a belief system.

Schull's approach is convenient, since it is nearly impossible to prove that leaders are sincere in their professed beliefs whereas one can measure consistency of action. However, this is nevertheless problematic since in the same way that the public statements of leaders expressing their ideological motivation cannot necessarily be taken at face value, the individual digressions from a professed belief system cannot necessarily serve as evidence of insincerity. Leaders may at times exploit ideologies for hidden purposes, and by the same token, they may compromise the personal beliefs they genuinely hold. Thus, if the term *ideology* becomes too distant from the concept of *individually held beliefs*, then another term can and should be employed to describe the phenomenon—as Schull himself recommends (and in his case the term recommended is discourse).[70] Hence throughout this article, the professed liberal economic beliefs of Klaus, Chubais, and others are understood as genuine, even if not always consistently applied to policy.

In this regard, ideological studies share some of the methodological challenges associated with cultural studies. Most notably both raise the necessity of identifying a given ideological context through rich descriptive interpretive analysis rather than through quantifiable, reproducible measurement. Despite this difficulty and the inherent subjectivity of doing so, the importance of ideology—even on an intuitive level—requires that the interpretive effort be made.

IDEOLOGY AND EXPLANATION

Given the importance of ideological factors for the development of property rights systems, why have they received so little attention in the property rights literature? In part, the answer to this question lies in the methodological foundation of property rights theory. Pioneers of property rights analysis such as Harold Demsetz and Douglass North based their theoretical approach in large part upon the techniques and the central assumptions of the dominant paradigm in economics, which emphasized the importance of material conditions and the rational calculation of interests for explaining material outcomes.[71] Although North currently acknowledges the centrality of ideology in the development of economic institutions, it is his earlier approach (in which ideas are justifications of economic interests) that inspired the future course of property rights analysis.[72]

[70] In fact, this is precisely what Schull (fn. 68) recommends, that the term "ideology" be replaced with the term "discourse" in instances where an ideas system is not actually "believed in" but constrains the policy debate nonetheless.

[71] Demsetz (fn. 5); and North (fn. 5).

[72] North (fn. 13).

Later analyses of property rights formation by political economists, such as Gary Libecap, William Riker, Itai Sened, and David Weimer, took the property rights literature in a new direction.[73] Incorporating political power and interests into their explanations, political economists developed models to show that property rights emerge from the efforts of key individuals to amass not only material resources but also political power. Yet by focusing on leaders' "utility maximization" (in general terms) rather than on their "wealth maximization" (in a narrow material sense), political economists extended the property rights literature without breaking with the foundations of its earlier contributions.

The lack of attention given to ideological factors is hardly unique to property rights theory. It is also characteristic of much of the literature in political economy because the dominant methodological approach is built upon rational choice foundations. In recent years, however, a group of political theorists has brought renewed interest to studying the generative power of ideas and to reconsidering the ideological and cognitive components of policy-making and regime change. This literature, at times referred to as the "ideas literature," focuses on areas as diverse as the introduction of the developmentalist model in Latin America, Keynesian economics in Western Europe, and the free trade regime in the United States, in order to develop generalizable theories about how ideas affect the introduction of new economic paradigms.[74] The contributors to this literature help to fill the gap left by mainstream social science theory, which, like property rights theory, rests upon rational choice foundation and thus typically offers a simplified rendering of preference formation and human rationality. The goal of the present article on property rights, like many other studies in the ideas literature, is to complement and further refine interest-based approaches by analyzing interest perception and interest specification as ideological events.[75]

[73] Riker and Sened (fn. 5); Riker and Weimer (fn. 5); Libecap (fn. 5); Eggertsson (fn. 5).

[74] The theorists included among the contributors to the ideas literature (e.g., Kathryn Sikkink, Peter Hall, Judith Goldstein, Robert Keohane, John Odell, G. John Ikenberry, Jeffrey Checkel) help to identify the social and structural mechanisms through which ideas as shared beliefs affect the creation and persistence of new policy projects, political paradigms, and regimes. In contrast to analytical approaches based upon cognitive psychology, the ideas of individuals are important in the ideas literature only inasmuch as they relate to a large community holding the same beliefs. See two review articles focusing on the ideas literature and related literatures: Jacobsen (fn. 57); and Albert Yee, "The Causal Effects of Ideas on Policies," *International Organization* 50 (Winter 1996).

[75] See Judith Goldstein, "The Impact of Ideas on Trade Policy: The Origins of U.S. Agricultural and Manufacturing Policies," *International Organization* 43 (Winter 1989), 32–34; Ikenberry (fn. 52), 59; Robert Jackson, "The Weight of Ideas in Decolonization: Normative Change and International Relations," in Goldstein and Keohane (fn. 52), 112–13.

A greater focus on ideas would enhance many other bodies of theory n political economy. For instance, an explicit treatment of ideas would lso contribute to the literature on structural adjustment—a literature hat examines the ability of governments to implement the liberal eco- omic policies promoted or required by international financial institu- ions such as the International Monetary Fund and the World Bank. ome attention is paid in this literature to the ideological orientation f economic reformers and the culture of the bureaucracy; the delegit- mization or discrediting of interest groups; and leaders' understanding f, and interaction with, transnational academic and cosmopolitan ad- isory communities.[76] Frequently, however, a greater emphasis is placed n nonideological structural factors, such as the insulation of tech- ocrats, or on links to powerful economic groups in society.[77]

Additionally, an inclusion of ideological factors would also enhance ur understanding of the implementation of political regimes, such as he human rights regime, as some contributors to the ideas literature ave already begun to explore.[78] Even the study of security regimes has ecently been enriched by considering nonmaterial ideological vari- bles.[79] Certainly, some of the parsimony of generalizable theories is ost by including nonmaterial variables. However, many of the gaps in ur understanding of political and economic phenomena, which be- ome especially apparent during comparative study, are narrowed hrough a greater appreciation of the role of ideology.

[76] On the culture of bureaucracy, see Byung Chol Koh, *Japan's Administrative Elite* (Berkeley: University of California Press, 1989); on "cosmopolitanness," see Thomas Callaghy, "Vision and Politics in the Transformation of the Global Political Economy: Lessons from the Second and Third Worlds," in Robert Slater, Barry Schultz, and Steven Dorr, eds., *Global Transformation and the Third World* (Boulder, Colo.: Lynne Rienner, 1993). On interest-group legitimacy, see Joan M. Nelson, "The Politics of Economic Transformation: Is Third World Experience Relevant in Eastern Europe?" *World Politics* 45 (April 1993).

[77] On the insulation of elites, see Stephen Haggard and Robert Kaufman, eds., *The Politics of Adjustment: International Constraints, Distributive Conflicts, and the State* (Princeton: Princeton University Press, 1992), introduction. On networks between the state and economic groups, see Peter Evans, "The State as Problem and Solution: Predation, Embedded Autonomy, and Structural Change," in Haggard and Kaufman.

[78] See Kathryn Sikkink, in Goldstein and Keohane (fn. 52), chap. 6.

[79] Peter Katzenstein, *Cultural Norms and National Security* (Ithaca, N.Y.: Cornell University Press, 1996).

Review Articles

WHAT'S SO DIFFERENT ABOUT A COUNTERFACTUAL?

By RICHARD NED LEBOW*

Niall Ferguson, ed. *Virtual History: Alternatives and Counterfactuals.* New York: Basic Books, 1999, 500 pp.
Niall Ferguson. *The Pity of War: Explaining World War I.* New York: Basic Books, 1999, 563 pp.

BECAUSE the siege of Troy was proving difficult, the frustrated Argives went down to their ships to sail for home. According to Homer, "A homecoming beyond fate might have been accomplished had not Hera spoken a word to Athene." The grey-eyed Athena sped down the peak of Olympus and instructed Odysseus to prevent the departure of the Argives, which he did.[1] The drama of the *Iliad* rests on a counterfactual, as do many other great works of literature. Counterfactuals are also taken seriously in the physical and biological sciences, where researchers routinely use them to develop and evaluate sophisticated, nonlinear models.[2] They have been used with telling effect in American politics, and other political scientists could greatly benefit from wider use of this useful research tool.[3] But for most members of

* I would like to thank Richard Hamilton, Richard Herrmann, Edward Ingram, Friedrich V. Kratochwill, Eli Bohmer Lebow, Geoffrey Parker, Janice Gross Stein, and Philip E. Tetlock for their helpful comments.

[1] *The Iliad of Homer*, trans. Richard Lattimore (Chicago: University of Chicago Press, 1951), bk. 2 lines 135–210.

[2] There are numerous examples from physics. On using information from nonevents to test nuclear weapons, see R. Penrose, *Shadows of the Mind: A Search for the Missing Science of Consciousness* (Oxford Oxford University Press, 1994); and A. C. Elitzur and L. Vaidman, "Quantum-Mechanical Interaction-Free Measurement," *Foundations of Physics* 23, no. 7 (1993), 987–97. On using interaction-free measurements with a test particle to determine the presence of an object, see G. Kwiat et al., "Interaction Free Measurements," *Physics Review Letters* 74, no. 12 (June 12, 1995), 4763–66. And on counterfactual computation, see Graeme Mitchison and Richard Josza, "Counterfactual Computation," *Quantum-Ph/9907007* (July 2, 1999).

[3] Electoral studies are a case in point. See Raymond E. Wolfinger and Benjamin Highton, "Can More Efficient Purging Boost Turnout" (Paper presented at the annual meeting of the American

World Politics 52 (July 2000), 550–85

our profession counterfactual arguments appear to have no scientific standing. They are flights of fancy, fun over a beer or two in the faculty club, but not the stuff of serious research.[4]

I begin my essay with the proposition that the difference between so-called factual and counterfactual arguments is greatly exaggerated; it is one of degree, not of kind. I go on to discuss three generic uses of counterfactual arguments and thought experiments. In the process, I distinguish between "miracle" and "plausible" world counterfactuals and identify the uses to which each is suited. I critique two recent historical works that make extensive use of counterfactuals and contend that they are seriously deficient in method and argument. I then review the criteria for counterfactual experimentation proposed by social scientists who have addressed this problem and find many of their criteria unrealistic and overly restrictive. The methods of counterfactual experimentation need to be commensurate with the purposes for which they are used, and I conclude by proposing eight criteria I believe appropriate to plausible-world counterfactuals.

COUNTERFACTUALS VERSUS FACTUALS

Counterfactuals are "what if" statements, usually about the past. Counterfactual experiments vary attributes of context or the presence or value of variables and analyze how these changes would have affected outcomes. In history and political science these outcomes are always uncertain because we can neither predict the future nor rerun the tape of history.

The speculative nature of counterfactuals makes many scholars wary of them. But counterfactual analysis is not always based solely on speculation. In the aftermath of the arrest of Aldrich Ames as a Soviet spy, the Central Intelligence Agency convened a team of counterintelli-

Political Science Association, New York, 1994); Ruy A. Teixeira, *The Disappearing American Voter* (Washington, D.C.: Brookings Institution, 1992); Steven J. Rosenstone and John Mark Hansen, *Mobilization, Participation, and Democracy in America* (New York: Macmillan, 1993); Richard J. Timpone, "Past History Reverses: Counterfactuals and the Impact of Registration Reform on Participation and Representation" (Manuscript, 1999).

[4] According to A. J. P. Taylor, "A historian should never deal in speculation about what did not happen." See Taylor, *The Struggle for the Mastery in Europe, 1848–1918* (London: Oxford University Press, 1954); E. H. Carr, *What Is History?* (London: Macmillan, 1961), 127; E. P. Thompson, *The Poverty of Theory and Other Essays* (New York: Monthly Review Press, 1978), 300. M. M. Postan writes: "The might-have beens of history are not a profitable subject of discussion"; quoted in J. D. Gould, "Hypothetical History," *Economic History Review*, 2d ser., 22 (August 1969), 195–207. See also David Hackett Fischer, *Historians' Fallacies* (New York: Harper Colophon Books, 1970), 15–21; Peter McClelland, *Causal Explanation and Model-Building in History, Economics, and the New Economic History* (Ithaca, N.Y.: Cornell University Press, 1975).

gence experts to figure out how he might have been unmasked earlie
The team imagined a series of procedures that might have been in pla
and asked which, if any of them, might have tripped up Ames. Usir
their extensive knowledge of Ames's personality, motives, and behavi
and of the modus operandi of his Soviet spymasters, they were able
conduct their inquiry with some degree of precision.[5] And in New Yo
City authorities only recently prepared for the possibility of widespre
Y2K computer failures by running tests of the ability of city agencies
respond to a complex emergency. Officials were confident that suc
breakdowns would not occur, but they used the opportunity to consid
the scenarios as "future counterfactuals" to gather useful data.[6] The co
troversy surrounding the strategy of deterrence provides an example
the use of counterfactuals in international relations. One of the princ
pal policy lessons of the 1930s was that appeasement whets the ap
petites of dictators while military capability and resolve restrains then
The failure of Anglo-French efforts to appease Hitler is well estal
lished, but the putative efficacy of deterrence rests on the counterfa
tual that Hitler *could* have been restrained *if* France and Britain ha
demonstrated willingness to go to war in defense of the European te
ritorial status quo. German documents make this an eminently re
searchable question, and historians have used these documents to try
determine at what point Hitler could no longer be deterred.[7] The
findings have important implications for the historical assessment
French and British policy and for the strategy of deterrence.

The Cuban missile crisis is an evidence-rich environment in whic
counterfactuals drove policy and have since been subjected to testin
Nikita Khrushchev's decision to send missiles to Cuba and then to re
move them and Kennedy's decision to impose a blockade we
contingent upon hypothetical antecedents. Kennedy believed—inco
rectly—that Khrushchev sent missiles to Cuba because he doubted th
president's resolve and that he would not have done so had Kenne
taken a stronger stand at the Bay of Pigs or in Berlin. Kennedy rea
soned that he had to compel Khrushchev to remove the missiles to co
vince him of his resolve and to deter a subsequent and more serio
Soviet challenge to the Western position in Berlin. Evidence from So
viet and American archives and interviews with former officials mak

[5] Private communication to the author.

[6] *New York Times*, November 27, 1999, A1.

[7] Yuen Foong Khong, "Confronting Hitler and Its Consequences," in Philip E. Tetlock and Aar
Belkin, eds., *Counterfactual Thought Experiments in World Politics: Logical, Methodological, and Psych
logical Perspectives* (Princeton: Princeton University Press, 1996).

it possible to explore the validity of most of these counterfactuals and thus to evaluate the choices of Soviet and American leaders and the subsequent scholarly analyses of the crisis.[8]

Quantitative counterfactual analysis is another possibility. Jay Winter exploits counterfactual projections of mortality rates based on prewar data from Prudential Life Insurance policies to determine the age structure of British war losses. He combines data from the life tables for 1913 and 1915 in roughly two to one proportions, as the war did not begin until August 1914, to create a counterfactual table for 1914. On the basis of the prior decade, he then calculates what the life tables would have been for the period 1914–18 in the absence of war. By comparing the actual death rates in each age group with the counterfactual estimates, he is able to determine the death rates of five-year cohorts for each year of the war.[9]

Even when evidence is meager or absent, the difference between counterfactual and "factual" history may still be marginal. Documents are rarely smoking guns that allow researchers to establish motives or causes beyond a reasonable doubt. Actors only occasionally leave evidence about their motives, and historians rarely accept such testimony at face value. More often historians infer motives from what they know about actors' personalities and goals, their past behavior, and the constraints under which they operated. In his highly acclaimed study of the Peloponnesian War, Donald Kagan argues that Pericles wanted to ally with Corcyra in the expectation that it would deter Sparta from coming to the aid of Corinth. If deterrence failed, Athens, protected by its city walls and the long walls to its harbor at Piraeus, would refuse to engage the main body of Spartan forces even if they invaded Attica and laid waste its olive groves and vineyards. After the Spartans experienced a few years of frustration, Pericles expected them to recognize the futility of waging war against Athens. Further, he thought that the peace faction, led by King Archidamus, would regain power and that the two hegemons would reach a more lasting accommodation.[10] The purpose of this purely speculative scenario is to explain away behavior that otherwise appears unenlightened and warlike.

The problem of motive is not unique to ancient history, where sources are notoriously meager. Janice Gross Stein and I spent several

[8] Richard Ned Lebow and Janice Gross Stein, "Back to the Past: Counterfactuals and the Cuban Missile Crisis," in Tetlock and Belkin (fn. 7).

[9] Winter, *The Great War and the British People* (London: Macmillan, 1986), 76–83.

[10] Kagan, *The Outbreak of the Peloponnesian War* (Ithaca, N.Y.: Cornell University Press, 1969), 203–342.

years researching a book on cold war crises. We scoured archives in four countries, utilized documents collected or declassified by other researchers, and conducted extensive interviews with former American, Soviet, Israeli, and Egyptian policymakers. We accumulated a mass of relevant information but still had no hard evidence about the motives for some of the key decisions made by Kennedy and Khrushchev. We suspect that Khrushchev was never clear in his own mind about the relative importance of the several goals that made a missile deployment in Cuba attractive to him. Given the delicate nature of many crisis decisions, neither leader was willing to share his goals and reasoning with even his most intimate advisers. Khrushchev further complicated the picture by telling various officials what they wanted to hear and thus what he thought was most likely to garner their support.[11]

When we move from the level of analysis of individual actors to small groups, elites, societies, states, and regional and international systems, the balance between evidence and inference shifts decisively in the direction of the latter. Structural arguments assume that behavior is a response to the constraints and opportunities generated by a set of domestic or international conditions. Mark Elvin's elegant study of China starts from the premise that empires expand to the point at which their technological superiority over their neighbors is approximately counterbalanced by the burdens of size. At this equilibrium, imperial social institutions come under constant strain because of the high relative cost of security. Harsh taxation impoverishes peasant cultivators and leads to falling tax revenues. The ensuing decline in the number of free subjects makes military recruitment more difficult, forcing governments to rely instead on barbarian auxiliaries, even for their main fighting forces. To save money, governments also give up active defense policies and try to keep hostile barbarians at bay through diplomacy, bribery, and settlement on imperial lands. The inevitable outcome is a weakened economic base, barbarization from within, and finally partial or total collapse of the empire.[12] Elvin musters considerable evidence in support of his thesis, much of it from primary sources, but it is all by way of illustration. Nowhere is he able to show that Chinese leaders adopted any of the policies he describes for any of the reasons he attributes to them.

For the most part, then, structural arguments are built on a chain of inference that uses behavioral "principles" as anchor points. Empirical evidence, when available, may be exploited to suggest links between

[11] Lebow and Stein, *We All Lost the Cold War* (Princeton: Princeton University Press, 1994).
[12] Elvin, *The Pattern of the Chinese Past: A Social and Economic Interpretation* (Stanford, Calif.: Stanford University Press, 1973).

hese principles and behavior. But even in the best of cases these links re indirect and presumptive and can be corroborated only obliquely nd incompletely. Readers evaluate these arguments on the seeming reasonableness" of the inferences drawn, the quality and relevance of he evidence offered in support, and the extent to which that evidence ermits or constrains alternative interpretations. Receptivity to arguments is significantly influenced by the appeal of the underlying political and behavioral "principles" in which the inferences are rooted. Vhen these "principles" run counter to the reigning orthodoxy, the aruments may be dismissed out of hand regardless of the evidence.

Good counterfactual thought experiments differ little from "factual" nodes of historical reconstruction. If we attempt to evaluate the imortance of Mikhail Gorbachev for the end of the cold war by considring the likely consequences of Chernenko being succeeded by omeone else, we need to study the career and policies of other possible uccessors (for example, Grishin, Romanov, Ligachev) and infer their olicies on the basis of their past preferences and commitments, the olitical environment in 1985, and the general domestic and foreign ituation of the Soviet Union.[13] There is a lot of evidence about all three uestions, evidence that sustains informed arguments about the kind of omestic and foreign policies these leaders might have pursued. Adnittedly, unexpected events, like Mathias Rust's Cessna flight to Red quare in May 1987, which Gorbachev exploited to purge the military f many hard-liners, can also have significant influence on policy.

The difference between factual and counterfactual arguments is furher blurred when we recognize that, as in the Cuban missile crisis, we ften need to understand the factual *and* counterfactual beliefs of hisorical actors to account for their behavior. In the missile crisis beliefs haped arguments: in the absence of compelling evidence, the beliefs of fficials determined the motives they attributed to Khrushchev for deloying Soviet missiles in Cuba, their estimates of the cost calculations nd political conflicts they assumed to being taking place in Moscow, nd the likely Soviet responses to a blockade, air strike, or invasion. ome of these beliefs took the form of conditional expectations, and vith the passage of time they became historical counterfactuals.

[13] See George Breslauer and Richard Ned Lebow, "Leadership and the End of the Cold War: A ounterfactual Thought Experiment," in Richard Herrmann and Richard Ned Lebow, eds., *Learning om the Cold War* (forthcoming). The authors identify leaders other than Gorbachev, Reagan, and ush who might have come to power in the Soviet Union and the United States and play out the relting interaction. They find that different strategies or tactics by either superpower or their allies ould have speeded up, slowed down, altered, or derailed the process of accommodation that led to the d of the cold war in 1990–91.

Counterfactuals are frequently smuggled into so-called factual narratives. E. H. Carr, no friend of counterfactuals, did this in his treatment of the Soviet Union when he insisted that the Bolshevik Revolution was highjacked by Stalin. The implication is that socialism would have developed differently without him.[14] After Cuba, former Kennedy administration officials and many scholars maintained that Khrushchev would not have deployed missiles in Cuba if Kennedy had been more decisive at the Bay of Pigs, at the Vienna summit, and in Berlin. There was no evidence to support this interpretation, but it became the conventional wisdom and helped to shape a host of subsequent policy decisions, including the disastrous intervention in Vietnam. The evidence that came to light in the Gorbachev era suggested, to the contrary, that Khrushchev decided to send missiles secretly to Cuba because he *overestimated* Kennedy's resolve. He feared that Kennedy, preparing to invade Cuba, would send the American navy to stop any ships carrying missiles to Cuba to deter that invasion.[15] Counterfactual arguments, like any historical argument, are only as compelling as the logic and "evidence" offered by the researcher to substantiate the links between the hypothesized antecedent and its expected consequences. Every good counterfactual thus rests on multiple factuals, just as every factual rests on counterfactual assumptions—and these assumptions too often go unexamined.

Any sharp distinction between factuals and counterfactuals rests on questionable ontological claims. Many of the scholars who dismiss counterfactual arguments do so because they do not believe they are based on facts. Philosophers have long recognized that "facts" are social constructions. They do not deny that reality exists quite independent of any attempt to understand it by human beings or that some understandings may transcend culture. Physical scientists may be correct in their claims that fundamental concepts like mass, volume, and temperature are essential to the study of nature and that extraterrestrial scientists would have to possess the same concepts to understand the universe. This is not true of social concepts, which vary across and within human cultures. There are many ways of describing social interactions, and the choice and utility of concepts depend largely on the purpose of the "knower."[16]

"Temperature" is undeniably a social construction, but it is a measure of something observable and real: changes in the energy levels of mol-

[14] Carr, *Socialism in One Country, 1924–1926* (New York: Macmillan, 1958–1964), 1:151.
[15] Lebow and Stein (fn. 11) review this argument and the new evidence (chap. 4).
[16] John R. Searle, *The Construction of Social Reality* (New York: Free Press, 1995).

ecules. Social and political concepts do not describe anything so concrete. There is no such thing as a balance of power, a social class, or a tolerant society. Social "facts" are reflections of the concepts we use to describe social reality, not of reality itself. They are ideational and subjective, and even the existence of "precise" measures for them—something we have only rarely—would not make them any less arbitrary. For, as Quine has shown, theoretical concepts insinuate themselves into the "data language" of even the hardest sciences.[17] The construction of "factual" history is therefore entirely imaginary, and its only claim to privilege is that the concepts and categories in terms of which it is constructed tell us something useful or interesting about the social world. The same is true for counterfactual history.

Counterfactuals can be used experimentally to substantiate Quine's claim that there is no conceptually neutral data language. Tetlock and Lebow asked a group of foreign policy experts to assess the contingency of the outcome of the Cuban missile crisis. One of their experiments used a "factual framing" of the question (at what point did some form of peaceful resolution of the Cuban missile crisis become inevitable?) and a counterfactual framing (at what point did all alternative, more violent outcomes become impossible?). From a logical point of view, the two questions are strictly complementary. Knowing the answer to either question, we should be able to deduce the answer to the other. Even though these two measures were obtained almost side by side in our questionnaire, the factual versus counterfactual framings of the historical question elicited systematically different responses, not just random variation that could be attributed to fatigue or boredom. Experts perceived substantially more contingency when they reflected on the counterfactually framed question. This is a good empirical demonstration of the importance of the benchmark against which the outcome is compared and offers support for the constructivist claim that how we pose "purely empirical" questions systematically shapes the answers we find.[18]

WHY COUNTERFACTUALS?

Counterfactuals can combat the deeply rooted human propensity to see the future as more contingent than the past, reveal contradictions in our

[17] Donald Davidson and Jaakko Hintikka, eds., *Words and Objection: Essays on the Work of W. V. Quine* Dordrecht: D. Reidel, 1969).

[18] Philip E. Tetlock, "Close-Call Counterfactuals and Belief System Defense: I Was Not Almost Wrong but I Was Almost Right," *Journal of Personality and Social Psychology* 75 (September 1998), 539–52.

belief systems, and highlight double standards in our moral judgments. Counterfactuals are an essential ingredient of scholarship. They help determine the research questions we deem important and the answers we find to them. They are also necessary to evaluate the political, economic, and moral benefits of real-world outcomes. These evaluations in turn help drive future research.

RECEPTIVITY TO CONTINGENCY

International relations theorists seek to understand the driving forces behind events; they usually do so after the fact, when the outcome is known. The process of backward reasoning tends to privilege theories that rely on a few key variables to account for the forces allegedly responsible for the outcomes in question. For the sake of theoretical parsimony, the discipline generally favors independent variables that are structural in nature (for example, balance of power, state structure, size and nature of a coalition). The theory-building endeavor has a strong bias toward deterministic explanations and on the whole downplays understandings of outcomes as the products of complex, conjunctional causality.[19] A recent survey of international relations specialists revealed that those scholars who were most inclined to accept the validity of theories (for example, power transition, nuclear deterrence) and theory building as a scholarly goal were the most emphatically dismissive of plausible-world counterfactuals. They were also most likely to invoke second-order counterfactuals to get developments diverted by counterfactuals back on the track.[20]

In retrospect, almost any outcome can be squared with any theory unless the theory is rigorously specified. The latter requirement is rarely met in the field of international relations, and its deleterious effect is readily observed in the ongoing debate over the end of the cold war. Various scholars, none of whose theories predicted a peaceful end to that conflict, now assert that this was a nearly inevitable corollary of their respective theories.[21] We observe a similar phenomenon in studies of Middle Eastern politics. Developments that seemed almost unthinkable before they happened—Sadat's trip to Jerusalem, the Pales-

[19] On conjunctional causality, see Charles C. Ragin, *The Comparative Method: Moving beyond Qualitative and Quantitative Strategies* (Berkeley: University of California Press, 1987).

[20] Tetlock and Lebow, "Poking Counterfactual Holes in Covering Laws: Alternative Histories of the Cuban Missile Crisis" (Manuscript, Mershon Center, Ohio State University, April 2000); on second-order counterfactuals, see Tetlock (fn. 18).

[21] William Wohlforth reviews this literature and criticizes realists for explaining ex post facto what none of them had predicted ex ante; see Wohlforth, "New Evidence on Moscow's Cold War: Ambiguity in Search of Theory," *Diplomatic History* 21 (Spring 1997).

tinian-Israeli moves toward peace—are subsequently described as having been overdetermined by structural causes, particularly shifts in the relative balance of power.[22]

The disciplinary tendency to privilege structural explanations is reinforced by the "certainty of hindsight bias."[23] Baruch Fischoff has demonstrated that "outcome knowledge" affects our understanding of the past by making it difficult for us to recall that we were once unsure about what was going to happen. Events deemed improbable by experts (for example, peace between Egypt and Israel, the end of the cold war) are often considered "overdetermined" and all but inevitable after they have occurred.[24] By tracing the path that appears to have led to a known outcome, we diminish our sensitivity to alternative paths and outcomes. We may fail to recognize the uncertainty under which actors operated and the possibility that they could have made different choices that might have led to different outcomes.

Many psychologists regard the certainty-of-hindsight effect as deeply rooted and difficult to overcome. But the experimental literature suggests that counterfactual intervention can assist people in retrieving and making explicit their massive but largely latent uncertainty about historical junctures, that is, to recognize that they once thought, perhaps correctly, that events could easily have taken a different turn. The proposed correctives use one cognitive bias to reduce the effect of another. Ross et al. exploited the tendency of people to inflate the perceived likelihood of vivid scenarios to make them more responsive to contingency. People who were presented with scenarios describing possible life histories of post-therapy patients evaluated these possibilities as more likely than did members of the control group who were not given the scenarios. This effect persisted even when all the participants in the experiment were told that the post-therapy scenarios were entirely hypothetical.[25] Philip E. Tetlock and the author conducted a

[22] Steven Weber, "Prediction and the Middle East Peace Process," *Security Studies* 6 (Summer 1997), 196.

[23] This point is also made by Philip E. Tetlock and Aaron Belkin, "Counterfactual Thought Experiments in World Politics: Logical, Methodological, and Psychological Perspectives," in Tetlock and Belkin (fn. 7), 15–16.

[24] Baruch Fischoff, "Hindsight Is Not Equal to Foresight: The Effect of Outcome Knowledge on Judgment under Uncertainty," *Journal of Experimental Psychology: Human Perception and Performance* 1, no. 2 (1975), 288–99; S. A. Hawkins and R. Hastie, "Hindsight: Biased Judgments of Past Events after the Outcomes Are Known," *Psychological Bulletin* 107, no. 3 (1990), 311–27. The tendency was earlier referred to as "retrospective determinism" in comparative-historical studies by Reinhard Bendix, *Nation-Building and Citizenship* (New York: Wiley, 1964).

[25] L. Ross et al., "Social Explanation and Social Expectation: Effects of Real and Hypothetical Explanations on Subjective Likelihood," *Journal of Personality and Social Psychology* 35 (November 1977), 817–29.

series of experiments to test the extent to which counterfactual unpacking leads foreign policy experts to upgrade the contingency of international crises. In the first experiment one group of experts was asked to assess the inevitability of the Cuban missile crisis. A second group was asked the same questions but was given three junctures at which the course of the crisis might have taken a different turn. A third group was given the same three junctures and, in addition, three arguments for why each of them was plausible. Judgments of contingency varied in proportion to the degree of counterfactual unpacking.[26] There is every reason to expect that scholars exposed to counterfactuals and, better yet, forced to grapple with their theoretical consequences will also become more open to the role of contingency in key decisions and events.

FRAMING RESEARCH

Research questions arise when events strike us as interesting or anomalous. To conceive of an event as anomalous we need a benchmark against which the outcome in question can be compared. Benchmarks can sometimes be derived from well-established laws or statistical generalizations: cold fusion, for example, would have been contrary to several of these laws and thus a truly anomalous event. There are few laws or statistical generalizations applicable to politics, but we hold to theories about how the political world works. They give rise to expectations, and when they are unfulfilled, to counterfactual worlds. These alternative worlds may appear more probable than the actual state of affairs. During the cold war the preeminent question in the security field was the long peace between the superpowers. In international political economy it was the survival of the postwar international economic order in the face of America's decline as a hegemon. Some security specialists considered it remarkable that the superpowers, unlike rival hegemons of the past, had avoided war. Some political economists were equally surprised that neither Germany nor Japan had sought to restructure international economic relations to their advantage in response to the apparent decline of the United States as hegemon.[27] Both

[26] The first of these experiments, involving alternative outcomes for the Cuban missile crisis, is described in Tetlock and Lebow (fn. 20).

[27] Kenneth N. Waltz, *Theory of International Politics* (Reading, Mass.: Addison-Wesley, 1979); John Lewis Gaddis, *The Long Peace: Inquiries into the History of the Cold War Era* (New York: Oxford University Press, 1987); Robert Gilpin, *War and Change in World Politics* (New York: Cambridge University Press, 1981), 232–33; Robert O. Keohane, *After Hegemony* (Princeton: Princeton University Press, 1984); Duncan Snidal, "The Limits of Hegemonic Stability Theory," *International Organization* 39 (Autumn 1985).

research agendas assumed that the status quo was an extraordinary anomaly that required an equally extraordinary explanation. For researchers who started from different premises—who assumed that none of the major powers were so unhappy with the current state of affairs that they were willing to risk war or economic disruption to change it—the seeming robustness of the political and economic orders posed no intellectual puzzle.

TESTING AND EVALUATION

Counterfactuals are fundamental to all theories and interpretations. If we hypothesize that x caused y, we assume that y would not have happened in the absence of x—ceteris paribus. Quantitative research attempts to get around this problem and around the contrapositive form of the fallacy of affirmation by constructing a sample of comparable cases large enough to contain adequate variation on dependent and independent variables.[28] James Fearon rightly observes that this strategy is effective only if there are no causes beyond those considered that vary systematically with the error term. To rule out this possibility, researchers need to pose the counterfactual of what would have happened if variables in the error term were altered. In actual experiments, this problem can only partially be solved by random assignment.[29] In case studies and historical narratives the problem is more pronounced because of the usual "loading up of explanatory variables."

Historians and case-study researchers typically attempt to establish causation by process tracing. They try to document the links between a stated cause and a given outcome in lieu of establishing a statistical correlation. This works best at the individual level of analysis but only when there is enough evidence to document the calculations and motives of actors. Even when such evidence is available, it may still not be possible to determine the relative weight of the several hypothesized causes and which of them, if any, might have produced the outcome in the absence of others or in combination with other causes not at work in the case. To sustain causal inference it is generally necessary to engage in comparative analysis. Within the single-case format comparative analysis can take two forms: intracase comparison and counterfactual analysis.

[28] Robyn M. Dawes, "Counterfactual Inferences as Instances of Statistical Inferences," in Tetlock and Belkin (fn. 7), 301–8. Strictly speaking it does not follow that if x then y; therefore, if not-x, then not-y, because factors other than x may also cause y.

[29] Fearon, "Causes and Counterfactuals in Social Science: Exploring an Analogy between Cellular Automata and Historical Processes," in Tetlock and Belkin (fn. 7).

... breaks down a case into a series ... that are treated as separate and independent ... analysis. Numerous studies of arms control and ... make use of this technique.[30] Like any form of ... comparison try to show as much variatio... (the explanandum) and independent (the ... This is sometimes more difficult to do than in c... comparisons. The independence of cases is also more proble... process and outcome of past decisions are likely to have c... influence on subsequent decisions about similar issues. Bu... case comparison confers a singular benefit: it builds variation w... fundamentally similar political and cultural context, controlling ... than intercase comparison for many factors that may be importa... are otherwise unrecognized. Unfortunately, not every case can be ... broken down into multiple decision points for purposes of compariso... When intracase comparison is impossible, variation can be gener... within a case by counterfactual experimentation. This latter strategy ... at the core of many simulations where variables are given a wide ra... of counterfactual values to determine the sensitivity of the outcome ... changes in one or more of them. Counterfactual simulation can identi... key variables and the range of values at which they will be have th... most impact on the outcome. Information obtained this way, especiall... if it has counterintuitive implications, can guide subsequent empirica... work intended to test the model or generate information necessary to make it a better representation of reality. Counterfactual simulation can test theories more directly. Thomas Cusack and Richard Stoll reviewed the realist literature to identify major shared assumptions among and the principal differences between contending interpretations. They formalized a model that would allow a computer-based simulation of some of these variants and focused their analysis on two issues of paramount importance to realists: state survival and system endurance. They found that the behavior realists prescribe as conducive to these ends (for example, maintaining large military forces, seeking to increase

[30] See, for example, Alexander L. George and Richard Smoke, *Deterrence in American Foreign Policy Theory and Practice* (New York: Columbia University Press, 1974); Alexander L. George and William E. Simmons, eds., *The Limits of Coercive Diplomacy* (Boulder, Colo.: Westview, 1994); Ted Hopf, *Peripheral Visions: Deterrence Theory and American Foreign Policy, 1965–1990* (Ann Arbor: University of Michigan Press, 1994); Robert Jervis and Jack Snyder, eds., *Dominoes and Bandwagons: Strategic Beliefs and Great Power Competition in the Eurasian Rimland* (New York: Oxford University Press, 1991); Thomas Risse-Kappen, *Cooperation among Democracies: The European Influence on U.S. Foreign Policy* (Princeton: Princeton University Press, 1995).

...ses, especially those aimed at balancing) did ... survival of individual units or the system as a ... man developed a more focused critique of re... the inevitable consequence of international anar... contends, then it follows, he reasoned, that global ... rarely emerge in real or counterfactual worlds. But ... that they appear with regularity in counterfactual ... ecially under conditions of defense dominance, the best ... sm.[32]

...ctual experiments and simulations can tease out the as...—often unarticulated—on which theories and historical in...ons rest.[33] Apologists for the Soviet system insist that ...sm would have evolved differently had Lenin lived longer or ...ceeded by someone other than Stalin.[34] Attempts to address ...stion have not resolved the controversy but have compelled his... to be more explicit about the underlying assumptions that guide ...ustain contending interpretations of Stalin and the nature of the ...munist Party and the Soviet state. Those assumptions have now ...me the focus of controversy, and scholars have looked for evidence ... which to evaluate them. This process has encouraged a more so...sticated historical debate.

Because every causal argument has its associated counterfactual, crit... have available to them two generic strategies. They can offer a dif...ent and more compelling theory or interpretation—far and away the most common strategy—or they can show that the outcome in ques...ion would have happened in the absence of the hypothesized causes. ...ohn Mueller's study of the cold war is a nice example of the second strategy. In contrast to the conventional wisdom that attributes the long peace between the superpowers to nuclear deterrence, Mueller argues ...hat Moscow and Washington were restrained by their general satisfac...tion with the status quo, by memories of World War II, and by the human, economic, and social costs of large-scale conventional warfare. He contends that the unheralded destructiveness of nuclear weapons

[31] Thomas R. Cusack and Richard J. Stoll, *Exploring Realpolitik: Probing International Relations Theory with Computer Simulation* (Boulder, Colo.: Lynne Rienner, 1990). Their evaluation was based on earlier work by Stuart A. Bremer and Michale Mihalka, "Machiavelli in Machina: Or Politics among Hexagons," in Karl W. Deutsch, ed., *Problems of Modeling* (Boston: Ballinger, 1977).

[32] Cederman, "Rerunning History: Counterfactual Simulation in World Politics," in Tetlock and Belkin (fn. 7).

[33] Fearon (fn. 29).

[34] George W. Breslauer discusses this literature in Breslauer, "Counterfactual Reasoning in Western Studies of Soviet Politics and Foreign Relations," in Tetlock and Belkin (fn. 7).

was redundant and possibly counterproductive.[35] *This Mueller strateg readily lends itself to simulation. Researchers can build a counterfactua world and look at how actors behave under a wide range of conditions including those that subtract putative causal factors.*

ASSESSING OUTCOMES

Counterfactuals are a key component of evaluation. Was the development of nuclear weapons a blessing or a curse for humankind? What about affirmative action, free trade, or the growing economic and political integration of Europe? Serious and thoughtful scholars can be found on all sides of these controversies. Their arguments share one thing in common: the use of counterfactual benchmarks—most often, implicitly—to assess the merits of real-world policies, outcomes, or trends. Proponents of nuclear weapons who claim that nuclear weapons had beneficial consequences during the cold war imagine a superpower war, or at least a higher probability of one, in the absence of nuclear deterrence. Some critics of nuclear weapons, like John Mueller, argue that self-deterrence based on memories of the horrors of conventional war would have kept the peace. Other critics contend that nuclear weapons sustained the cold war and that it would have been less intense and possibly resolved earlier in their absence.

Assessment can be significantly influenced, or even determined, by the choice of counterfactual. The conventional wisdom holds that the allied victory in World War I was a good thing: it prevented an expansionist, continental power from achieving hegemony in continental Europe. This assessment represents the view of the world from the corporate boardrooms and corridors of power in London, New York, and Washington. From the perspective of, say, Polish Jewry, the outcome was a disaster. If Germany had won, there almost certainly would have been no Hitler and no Holocaust. In this case, the choice of counterfactual reflects the different interests of the various groups. As with historical analogies, the interesting and eminently researchable question becomes the extent to which counterfactuals guide evaluation or are chosen to justify positions that people have reached for quite different reasons.[36]

[35] John Mueller, *Retreat from Doomsday: The Obsolescence of Major War* (New York: Basic Books, 1989). And see the debate on this subject between Mueller, "The Essential Irrelevance of Nuclear Weapons: Stability in the Postwar World," and Robert Jervis, "The Political Effects of Nuclear Weapons: A Comment," both in *International Security* 13 (Fall 1988).

[36] For a strong statement of the former position with regard to historical analogies, see Yuen Foong Khong, *Analogies at War: Korea, Munich, Dien Bien Phu, and the Vietnam Decisions of 1965* (Princeton: Princeton University Press, 1992).

MUST COUNTERFACTUALS BE REALISTIC?

The several uses of counterfactuals I have described use "plausible" and "miracle" world counterfactuals. Plausible-world counterfactuals are intended to impress readers as realistic; they cannot violate their understanding of what was technologically, culturally, temporally, or otherwise possible. In a recent study of the origins of World War I, I imagine a world in which Archduke Franz Ferdinand and his wife, Countess Sophie, returned alive from their visit to Sarajevo.[37] This counterfactual strikes me as eminently plausible because their assassination was such a near thing. Princip's accomplice missed the royals en route to city hall, and Princip was lamenting his failure when the touring car carrying Franz Ferdinand and his wife came to a stop in front of the bar from which Princip had just emerged to allow the cars at the head of the procession to back up because they had made a wrong turn. Princip stepped forward and fired two shots at point-blank range into their touring car. With only a minimal rewrite of history—the procession is halted after the first attempt or subsequently stays on the planned route—the assassination could easily have been averted.

There are many plausible counterfactuals—historical near misses, if you like—that might have come to pass but probably would not have had any significance for the outcomes in question. Plausible counterfactuals must meet a second test: they must have a real probability of leading to the outcome the researcher intends to bring about. To demonstrate this, the researcher must construct a logical path between the counterfactual change and the hypothesized outcome, and meet other tests that are described in the last section of the review. Plausible-world counterfactuals are thought by some researchers to be the only legitimate kind of counterfactual.[38]

Miracle counterfactuals violate our understanding of what is plausible or even possible.[39] Take the following assertion: "If Bosnians had been blue-nosed dolphins, NATO would not have allowed their slaughter."[40] A landlocked province of aquatic mammals with ethnic identities would be a stretch even for Dr. Seuss—but it is morally provocative. It implies that NATO's belated intervention was somehow related to the

[37] Richard Ned Lebow, "Contingency, Catalysts and System Change," *Political Science Quarterly* (forthcoming).

[38] Jon Elster, *Logic and Society: Contradictions and Possible Worlds* (New York: John Wiley, 1978); G. Hawthorn, *Plausible Worlds: Possibility and Understanding in History and the Social Sciences* (New York: Cambridge University Press, 1991), 31–60.

[39] The term "miracle counterfactual" was coined by Fearon (fn. 29), 60.

[40] Tetlock and Belkin (fn. 23), 14.

Muslim affiliation of much of Bosnia's population and would come sooner if Bosnians had characteristics that evoked more symp in the West. Miracle counterfactuals are useful for purposes of th building and testing. I could hypothesize that Europe achieved its r itary advantage because it was the only region of the world where long-standing hegemony was established and because the result prolonged competition among its leading political units made th lean and mean, better armed, and more efficient in the use of larg scale violence.[41] To advance this hypothesis I must consider the cou terfactual of a hegemonic Europe—perhaps achieved by a bett organized and more astutely led Spain in the sixteenth century. To su tain the hypothesis I need to argue what a hegemonic Europe wou have been like, and how it would have differed from the historical Eu rope. Miracle counterfactuals are particularly useful in evaluating exis ing interpretations.

Mueller's world without nuclear weapons is a miracle counterfactu because it would require a massive rewrite of a century of scientific an political history to uninvent nuclear weapons, although the timing o their development might be altered by plausible-world counterfactuals The value of miracle counterfactuals derives not from their realism bu from the analytical utility of considering alternative worlds. Counterfactuals of this kind have a distinguished lineage. Euclid used one to prove that there are an infinite number of prime numbers, and Newton, to demonstrate that the universe could not be infinite with regularly distributed and fixed stars. If these conditions held, Newton argued, the sky would not be blue.[42]

Miracle counterfactuals are critical to assessment. To evaluate the relative merits of court-contested versus mediated divorces, we need to know something about the financial consequences of each mode of settlement for women. To do this, we can compare outcomes in states that encourage and discourage mediation in comparable samples of divorced couples.[43] But we confront a serious data problem if we are interested in ascertaining whether making divorce more difficult to

[41] Ludwig Dehio, *The Precarious Balance: Four Centuries of the European Power Struggle* (New York: Knopf, 1995). Dehio argues that competition among many independent units produced "fertile friction" among Greek city-states and in modern Europe.

[42] Newton reasoned that the energy reaching us from an individual star is E/r^2, where r is the distance of the star, and E is the average energy radiated by each star, and if the density (d) of stars in the universe is constant, the number of stars would be d r^3. The total energy produced by these stars would grow in a linear fashion with r, and rise to infinity in an infinite universe—neither we nor the earth would exist. Hence, the density of stars must decrease or the universe must be finite.

[43] Carol Bohmer and Marilyn Ray, "Effects of Different Dispute Resolution Mechanisms on Women and Children after Divorce," *Family Law Quarterly* 28, no. 2 (1994).

obtain would hold families together, as many conservatives allege. Legislation to make divorce incrementally more difficult has been debated in several states but rejected in all of them. Really tough divorce laws nationwide are unrealistic in the current social climate, and the abolition of divorce is even more improbable.[44] But miracle counterfactuals that use either premise as the starting point for research might generate interesting results that could raise the level of the policy debate. This is not dissimilar to the kind of research economists do all the time; they raise or lower prices of commodities well beyond any realistic market expectations to test consumer preferences.

HISTORICAL COUNTERFACTUALS

Recent publications have sparked renewed interest in counterfactuals. In political science the principal catalysts have been Fearon's 1991 article in this journal and Tetlock and Belkin's edited volume, *Counterfactual Experiments in World Politics*.[45] The earlier publication of Mueller's *Retreat from Doomsday* also drew a lot of attention, but it was focused more on his controversial conclusion than on his counterfactual method.[46] The two works under review have revived interest in counterfactuals in the discipline of history.

Virtual History does not make a good case for counterfactuals. Ferguson's long introduction is a literature review without much purpose. It offers no reasons for engaging in counterfactual history other than sensitizing readers to contingency, and it only briefly addresses the methods by which counterfactual experiments should be conducted. Ferguson criticizes earlier counterfactual works for inferring momentous consequences from "simple, often trivial change[s]." With undisguised scorn, he cites as an example Pascal's intentionally provocative counterfactual about Cleopatra's nose: if it had been ugly, Anthony would not have fallen for her, and the history of Rome might have been different (pp. 11–12). But what is wrong with small changes having big effects? Could anyone seriously doubt that the course of history would have been different if Pharaoh's daughter had not found a child in a basket in the reeds, if the Mongol fleet had not encountered a destructive typhoon en route to Japan, if the duke of Alba had not fallen sick

[44] Deborah Rhode, *Speaking of Sex: The Denial of Gender Inequality* (Cambridge: Harvard University Press, 1997), 184.

[45] James D. Fearon, "Counterfactuals and Hypothesis Testing in Political Science," *World Politics* 43 (January 1991); Tetlock and Belkin (fn. 7). I do not review the latter volume because I am coauthor of one of its chapters.

[46] See fn. 35 for Mueller and the follow-up debate on nuclear weapons between Mueller and Jervis.

in 1572, or if Hitler had died in trenches during World War I or near fatal traffic accident he suffered in the summer of 1930? Whe duke of Alba took to his bed, his inexperienced and arrogant son command of the forces laying siege to Haarlem, rejected the to offer of surrender on terms, and prolonged the Dutch rebellion aga Spanish rule. The sustained ability of the Dutch to resist infuria Philip II and his nephew, Alexander Farnese, duke of Parma. Tl convinced themselves that the Dutch resisted only because of Engl support and decided to deal with England directly. Geoffrey Parl suggests that the subsequent defeat of the Spanish Armada laid t American continent open to invasion and colonization by Northe Europeans and made possible the founding of the United States.[47] F want of aspirin, a continent may have been lost.

With reason, Max Weber insisted that the most plausible countei factuals were those that made only "minimal rewrites" of history.[48] Sup pose we want to evaluate Ronald Reagan's role in ending the cold wa by considering the likely course of that conflict in the late 1980s if h had not been president. It is more plausible to assume that Hinckley' bullet lodged a few millimeters closer to a vital organ than to concoct a complicated, multistep scenario to deprive Reagan of his 1980 electora triumph. As a general rule, the fewer and more trivial the changes we introduce in history, the fewer the steps linking them to the hypothesized consequent and the less temporal distance between antecedent and consequent, the more plausible the counterfactual becomes. Not every small change will have significant, longer-term consequences; many, perhaps most, changes are likely to have consequences that are dampened over time. The real problem of counterfactual thought experimentation is trying to determine which minimal rewrites will affect the course of history.

Let us return to Cleopatra's nose. An ugly proboscis is a small change that might have had a big effect. It could have dampened Anthony's ardor, with important consequences for the struggle for power among Caesars' successors. So too might Roman history have been different if Anthony had been gay. We do not object to such counterfactuals because they are trivial but because they are arbitrary and

[47] Henry Ashby Turner, Jr., *Geissel des Jahrhunderts: Hitler and seine Hinterlassenschaft* (Berlin: Siedler, 1989); Geoffrey Parker, *The Grand Strategy of Philip II* (New Haven: Yale University Press, 1998a), 129–34; idem, "The Repulse of the English Fireships," in Robert Cowley, ed., *What If?* (New York: G. P. Putnam's Sons, 1998b).

[48] Weber, "Objective Possibility and Adequate Causation in Historical Explanation," in *The Methodology of the Social Sciences* (1905; Glencoe, Ill.: Free Press of Glencoe, 1949).

ontrived. There is no particular reason why Cleopatra should have been less attractive or why Anthony should have had a different sexual orientation. Nor is it clear how these changes would have been brought about. Good counterfactuals arise from the context, and there must be compelling mechanisms to bring them into being.

Ferguson wants to legitimate counterfactual research, but his efforts to do so would put it in a straitjacket. He insists that we consider only those counterfactual scenarios that contemporary actors considered and committed to paper or some other form of record that is accepted by historians as a valid source.[49] Ferguson's criteria would exclude entire categories of plausible-world counterfactuals. It would limit counterfactuals to elites who made written records, to self-conscious decisions in which alternatives are likely to be carefully considered, and to political systems in which leaders and other important actors feel secure enough to write down their thoughts or share them with colleagues, journalists, family members, or friends. It would rule out all counterfactuals that were the result of impulsive behavior (or the lack of it), of human accident, oversight, obtuseness, or unanticipated error, of acts of nature, or of the confluence (or lack of it), or of independent chains of causation. We could not contemplate a world in which the duke of Alba remained healthy and did not relinquish command of his army to his son, Franz Ferdinand's touring car adhered to its planned route, Hitler died in an automobile accident, or Hinckley assassinated Ronald Reagan. Ferguson's criteria would also rule out all miracle counterfactuals.

The nine historical chapters of *Virtual History* are organized chronologically. They consist of one chapter each on the seventeenth and eighteenth centuries—England without Oliver Cromwell and America without a revolution—and seven on the twentieth century, ranging from an Irish Britain following the enactment of Home Rule in 1912 to the survival of the Soviet Union and Warsaw Pact in the absence of Mikhail Gorbachev. Some of the chapters are imaginative and convincingly argued (for example, John Adamson on England without Cromwell and J. C. D. Clark on British America) and sensibly ignore Ferguson's restrictive criteria for plausible-world counterfactuals. Other chapters are less convincing, among them, Ferguson's chapter on 1914. One chapter is downright nasty, partisan, and inappropriate to the volume: Diane Kunz does a hatchet job on John F. Kennedy without any serious counterfactual analysis. Ferguson's conclusion is as rambling as

[49] A similarly restrictive definition is offered by P. Nash, "The Use of Counterfactuals in History: A Look at the Literature," *SHAFR Newsletter* (March 1991).

his introduction and makes no attempt to tie the historical chapters together or draw any conclusions from their counterfactual explorations. Neither do any of the contributors in the absence of any guidance from the editor.

The Pity of War is a big book about big questions — ten of them to be precise. Ferguson wants to know if World War I was inevitable, why Germany's leaders gambled on war in 1914, why Britain intervened, whether the war was really popular at the outset, the extent to which propaganda kept it going, why the Schlieffen Plan failed, why Britain's economic advantages did not result in an earlier victory, why men kept fighting, why they stopped fighting, and who won the peace. To answer these questions, Ferguson relies heavily on counterfactual arguments on the grounds that they help recapture the uncertainty of the actors, to whom the future was merely a set of possibilities, and that they allow us to assess the extent to which they made the right choices.

Ferguson's questions are good ones; his answers are not. He appears to be following the A. J. P. Taylor approach to the writing of history: make outrageous arguments that stand history on its head, infuriate the critics, gain publicity, and with any luck, sell a lot of books. Taylor's *Origins of the Second World War* was a masterpiece of the genre.[50] It tried to shift much of the burden for the outbreak of war in 1939 from Adolf Hitler and Germany to Neville Chamberlain and Britain. Ferguson attempts to do the same—much less cleverly—for World War I.

Most historians assign primary responsibility for the First World War to Austria-Hungary and Germany. The dominant interpretation holds that Austria-Hungary exploited the assassination of Franz Ferdinand as the pretext for war with Serbia, and that Germany encouraged—even pushed—Austria toward decisive action.[51] Students of

[50] A. J. P. Taylor, *The Origins of the Second World War* (London: Hamish Hamilton, 1961).

[51] Fritz Stern, "Bethmann Hollweg and the War: The Limits of Responsibility," in Fritz Stern and Leonard Krieger, eds., *The Responsibility of Power: Historical Essays in Honor of Hajo Holborn* (Garden City, N.Y.: Doubleday, 1967); Konrad J. Jarausch, "The Illusion of Limited War: Chancellor Bethmann Hollweg's Calculated Risk, July 1914," *Central European History* 2, no. 1 (1969); idem, *The Enigmatic Chancellor: Bethmann Hollweg and the Hubris of Imperial Germany* (New Haven: Yale University Press, 1973); Karl Dietrich Erdmann, "War Guilt 1914 Reconsidered: A Balance of New Research," in H. W. Koch, ed., *The Origins of the First World War*, 2d ed. (Oxford: Oxford University Press, 1984); Wolfgang J. Mommsen, *Das Zeitalter des Imperialismus* (Frankfurt: Fischer Bucherei, 1969); idem, "Domestic Factors in German Foreign Policy before 1914," *Central European History* 6, no. 1 (1972); Andreas Hilgruber, "Riezlers Theorie des kalkulieren Risikos und Bethmann Hollwegs politische Konzeption in der Julikrise 1914," *Historische Zeitschrift* 202 (April 1966); Egmont Zechlin, *Krieg und Kriegrisiko: Zur deutschen Politik im Ersten Weltkrieg* (Düsseldorf: Droste, 1979); David Kaiser, "Germany and the Origins of the First World War," *Journal of Modern History* 55 (December 1983); Hartmut Pogge von Strandmann, "Germany and the Coming of War," in R. J. Evans and Hartmut Pogge von Strandmann, *The Coming of the First World War* (Oxford: Oxford University Press, 1988); Holger H. Herwig, *The First World War: Germany and Austria-Hungary, 1914–1918* (London: Arnold, 1998).

Austria-Hungary argue that its leaders acted out a combination of closely related foreign and domestic concerns. Since the publication of Kurt Riezler's diaries, something of a consensus has emerged among German historians that Chancellor Theobald von Bethmann-Hollweg did not seek to provoke a European war but recognized that an Austrian conflict with Serbia would be difficult to contain. He also believed that a continental war was inevitable and that Germany's chance of winning it would only diminish with every passing year.[52] Hence, sooner was preferable to later.

To the extent that Britain has been criticized for its policy in 1914, it has usually been for Sir Edward Grey's naïveté. The foreign secretary recognized only belatedly that Germany was playing a double game, and for most of the crisis he had tried to get Berlin to restrain Austria and cosponsor a conference to resolve the problem by diplomacy. It is sometimes alleged that Britain might have prevented war if only Grey had put Germany on notice early on in the crisis that Britain would come to the aid of France if Germany violated Belgian neutrality.[53] Ferguson propounds a different idea in *The Pity of War*. Germany was on the defensive and had been put there by other powers, most notably Britain: "The Committee of Imperial Defence meeting of 23 August (rather than the notorious meeting between the Kaiser and his military chiefs sixteen months later) was the real 'war council' which set the course for a military confrontation between Britain and Germany" (pp. 64–65). This meeting acquiesced in the strategy of sending an expeditionary force to the Continent if Britain sided with France in a war with Germany. Britain's plans for an aggressive land and naval policy, its naval agreement with France in the Mediterranean, and secret discussions with Russia in 1914 made "German fears of encirclement seem less like paranoia than realism" (p. 68).

Ferguson contends that the British rode roughshod over Germany's interests because "they realized she did *not* pose a threat." The Germans

---

chap. 1; John C. G. Röhl, "Germany," in Keith Wilson, ed., *Decisions for War, 1914* (New York: St. Martin's Press, 1995). On Austria, see Samuel R. Williamson, Jr., "Influence, Power, and the Policy Process: The Case of Franz Ferdinand," *Historical Journal* 17, no. 2 (1974); idem, "The Origins of World War I," *Journal of Interdisciplinary History* 18 (Spring 1988); idem, *Austria-Hungary and the Coming of the First World War* (London: Macmillan, 1990); and R. J. Evans, "The Habsburg Monarchy and the Coming of War," in Evans and Pogge von Strandmann, eds., *The Coming of the First World War*.

[52] Riezler, *Tagebücher, Aufsätze, Dokumente: Eingeleidet und hrsg. Von Karl Dietrich Erdmann* (Göttingen: Vandenhoeck & Ruprecht, 1972); Koch (fn. 51), introduction.

[53] See, for example, Sean M. Lynn-Jones, "Détente and Deterrence: Anglo-German Relations, 1911–1914," *International Security* 11 (Fall 1986). German scholarship is very skeptical about such claims. Karl Dietrich Erdmann, "War Guilt 1914 Reconsidered: A Balance of New Research," in Koch (fn. 51); Zechlin (fn. 51); Jarausch (fn. 51, 1973), 167, 170.

were not aggressive but were pushed toward war by the mili
rations of their neighbors, including Britain, and resolved to
they still had a chance to win (pp. 151–52). This argument l
paradoxical conclusion—and the key counterfactual of the l
"a *more* militaristic Germany might have averted the First V
(p. 142). The counterfactual rests on three assumptions: (1)
nancial constraint on Germany's military capability was "p
crucial factor in the calculations of the German General Sta
(2) that General Erich von Ludendorff's 1913 "maximum p
three hundred thousand men to the army was financially fe
(3) that a more powerful Germany would have been more re
its foreign policy (pp. 141–42).

The German general staff was deeply concerned about th
implication of Russian and French military preparations, b
because of the infamous Schlieffen Plan. It committed Gern
all-out offensive in the West to defeat France before the m
mobilizing Russia could bring its forces to bear against Gern
East. If the general staff had adopted a defensive strategy, a
propriate strategy for an allegedly status quo power, they cou
garded French and Russian military preparations w
equanimity. The Ludendorff plan may have been financially
the reasons Ferguson adduces: additional taxes could have b
or the Reichsbank, which was hoarding gold, could have p
substantial issue of treasury bills. But the counterfactual viol
the criteria Ferguson proposed in *Virtual History:* the Germa
lor never seriously considered the Ludendorff plan as an opti
not, because increased military expenditures in 1913 were, in
own words, "politically impossible" (p. 141). This admissio
rug out from under the counterfactual. The pièce de résista
guson's bald assertion that a more powerful Germany would
a more peaceful Germany—while providing no grounds wha
believing this could have been. It seems at least as likely t
larger army in 1914, German leaders would have been *more*
of victory and *more* intent on going to war. In this connectio
to remember that an important catalyst for the German deci
war in the July crisis was the perceived need to bolste
Hungary, and this consideration would still have been p
equally important.

Ferguson fails to consider any additional counterfactuals
have followed from his counterfactual. If Germany had in
armed forces, how would France and Russia have respond

had each increased their armed forces by only half as much, the German general staff would have been right back where it started. Expectation of tit-for-tat escalation—a characteristic of prewar arms races—may well have been an additional reason why Bethmann-Hollweg did not greet Ludendorff's proposal with enthusiasm. Any number of other changes are possible as a consequence of Ferguson's counterfactuals, and a good historian should attempt to identify the most important and explore their implications for Germany policy and the likelihood of war.

Ferguson's counterfactual is politically and methodologically unsound. Most of his other counterfactuals are similarly incomplete and unconvincing. Consider the alternative strategies by which Germany might have achieved a military victory. Ferguson rejects out of hand the *Ostaufmarsch* plan, by which the German offensive would have been directed against Russia instead of France, and the related suggestion that Falkenhayn should have refrained from attacking Verdun, remained on the defensive in the West, and concentrated on defeating Russia. He suggests that Ludendorff might have obtained a negotiated peace if he had offered to relinquish Belgium after his spring 1918 offensive stalled on April 5. Instead of pressing on with the attack, Ludendorff should have withdrawn German forces to the Hindenburg Line and asked the allies for an armistice. This is not a course of action that seems to have crossed Ludendorff's mind at the time, and Ferguson gives no reasons for believing that the allies would have responded favorably (pp. 315–17).

The mother of all World War I counterfactuals is Ferguson's outlandish suggestion that had Britain stood aside in 1914, even for a matter of weeks, "continental Europe could therefore have been transformed into something not wholly unlike the European Union we know today" (pp. 443–44, 457–61). This benign outcome assumes that Germany would have defeated France without the rapid intervention of the British Expeditionary Force—an argument about which Ferguson is only lukewarm elsewhere in the volume. It also assumes a long string of second-order counterfactuals (for example, Russia is also defeated, Berlin organizes the Continent in a manner that allows other nations to benefit, these states and people become reconciled to German dominance, Germany is a sated power and conducts no further wars in Europe and makes no additional colonial demands, and its political culture and institutions evolve to become something akin to the liberal, peaceful, Europe-oriented country that it is today). Ferguson does not address these questions and offers no chain of logic to connect

the counterfactual of a German victory to the peaceful and consensual Europe he posits as the long-term outcome. Ferguson is reacting to the arguments of Grey and "other Germanophobes" that a German victory would have been disastrous for British interests and goes full tilt—and foolishly so—in the opposite direction (pp. 166–68).

"Plausible World" Counterfactuals

More thoughtful historians and social scientists have pondered the problem of plausible-world counterfactuals and appropriate criteria for using them. There is no consensus about what constitutes a good counterfactual, but there is a common recognition that it is extraordinarily difficult to construct a robust counterfactual—one whose antecedent we can assert with confidence could have led to the hypothesized consequent. There are three reasons for this well-warranted pessimism: the statistical improbability of multistep counterfactuals, the interconnectedness of events, and the unpredictable effects of second-order counterfactuals.

Compound Probability

The probability of a consequent is a multiple of the probability of each counterfactual linking the hypothesized antecedent to it. Suppose I contend that neither world war nor the Holocaust would have occurred if Mozart had lived to the age of sixty-five.[54] Having pushed classical form as far as it could go in the Jupiter Symphony, his last three operas, and the requiem, Mozart's next dramatic works would have been the precursors of a new, "postclassicist" style. He would have created a viable alternative to romanticism that would have been widely imitated by composers, writers, and artists. Postclassicism would have kept the political ideas of the Enlightenment alive and held romanticism in check. Nationalism would have been more restrained, and thus Austria-Hungary and Germany would have undergone very different political evolution. This alternative and vastly preferable world has at least five counterfactual steps linking antecedent to consequent: Mozart must survive to old age and develop a new style of artistic expression; subsequent composers, artists, and writers must imitate and elaborate it; romanticism must become to some degree marginalized; and artistic developments must have important political ramifications. This last

[54] Richard Ned Lebow, "What If Mozart Had Died at Your Age?" (Manuscript, Mershon Center, Ohio State University, April 2000).

counterfactual presupposes numerous other enabling counterfactuals about the nature of the political changes that will lead to the hypothesized consequent (for example, internal reforms that resolve or reduce the threat of internal dissolution of Austria-Hungary, German unification under different terms, or at least a Germany satisfied with the status quo, no First World War, no Hitler and no Holocaust without Germany's defeat in World War I). Even if every one of this long string of counterfactuals had a probability of at least 50 percent, the overall probability of the consequent would be a mere .03 for five steps and a frighteningly low .003 for eight steps. This particular counterfactual may appear far-fetched, but most interesting counterfactuals are no less improbable statistically. They may start with a tiny and plausible alteration of the real world but then infer numerous follow-on developments to end up with a major change in reality.

INTERCONNECTEDNESS

Scholars not infrequently assume that one aspect of the past can be changed and everything else kept constant. Mueller's cold war counterfactual is a case in point. He analyzes postwar history as it actually happened, including the Cuban missile crisis, to see if any major outcome would have been different in the absence of nuclear weapons. But what incentive would Khrushchev have had to deploy conventionally armed missiles in Cuba? The missiles could not have deterred an American invasion and might well have invited one, and this was the very event Khrushchev hoped to prevent.[55] Without a Cuban missile crisis, which had significant consequences for the future course of Soviet-American relations, the cold war would have evolved differently, and the course of future superpower relations, benign or malign, is impossible to predict.

"Surgical" counterfactuals are no more realistic than surgical air strikes. Causes are interdependent and have important interaction effects. Even minimal rewrites of history may alter the context in such a way as to render the consequent moot or to undercut the chain of events or logic leading to it. Consider another missile crisis counterfactual: if Richard Nixon had won the 1960 presidential election—and he lost by the narrowest of margins—he would have ordered an air strike against the Soviet missiles in Cuba. It is reasonable to assume that Nixon would have preferred an air strike to a blockade because he was more hawkish than Kennedy and would not have had a secretary of defense like Robert McNamara to make the case for restraint. But for

[55] For this criticism of Mueller, see Cederman (fn. 32), 253, 255.

these same reasons, Nixon might well have committed American forces to the faltering Bay of Pigs invasion in April 1961. If Castro had been overthrown, there would have been no communist Cuba to which Khrushchev could send missiles a year later.

The Nixon example invokes a counterfactual arising from the antecedent—but one outside the chain of logic leading from it to the consequent—to render the missile crisis moot. If Nixon had been elected president in 1960, the world would have been different in many ways, some of them with implications that are impossible to trace. He would have appointed a different defense secretary, who in turn would have appointed a different chairman of the joint chiefs of staff. Personnel changes at the top would have had amplifying consequences for promotions and appointments further down the line. To the extent that the behavior of individual officers could have important foreign policy implications—and it certainly did during the Berlin and Cuban missile crises—all kinds of possibilities open up.

SECOND-ORDER COUNTERFACTUALS

The problem of prediction is further complicated by the fact that the clock of history does not stop if and when the hypothesized consequent is reached. Subsequent developments can return history to the course from which the antecedent was intended to divert it.[56] Colin Martin and Geoffrey Parker show that the defeat of the Spanish Armada was a near event; they suggest that better communication, different decisions by local commanders, or better weather might have allowed the Spanish to land an invasion force in England. If Spain had put an army ashore, it almost certainly would have conquered the country. Martin and Parker go on to consider what would have happened next: Philip II was succeeded by Philip III, a far less capable ruler, who would have had enormous difficulty in maintaining an already overextended empire. In relatively short order, they believe, England would have overthrown the Spanish yoke.[57]

Some counterfactuals, like the "butterfly effect," introduce small changes that have major, lasting, long-term effects. Others, like a Spanish victory in 1588, are big changes that appear to have big consequences, but the changes they introduce may be damped down over time and end up having little lasting effect. It is also possible that second counterfactuals arising from a Spanish occupation of England,

[56] Tetlock, "'Distinguishing Frivolous from Serious Counterfactuals" (Manuscript, 1999).
[57] Parker (fn. 47, 1998a), 281–96.

however brief that occupation, could have had dramatic and long-lasting changes for European politics or in other realms that Marin and Parker did not consider. None of these outcomes is predictable, and the butterfly effect may not be knowable in advance or even in retrospect.[58]

CRITERIA FOR USE

Recognition that counterfactual arguments often have indeterminate consequences has prompted scholars to impose restrictive criteria on their use. Fearon proposes a proximity criterion. We should consider only those counterfactuals in which the antecedent appears likely to bring about the intended consequent and little else. Counterfactuals, he suggests, must be limited to cases where "the proposed causes are temporally and, in some sense, spatially quite close to the consequents."[59] Dawes argues that counterfactual inferences are warranted "if and only if they are embedded in a system of statistical contingency for which we have reasonable evidence."[60] Kiser and Levi suggest that counterfactuals are best used as substitutes for direct empirical analysis when data are limited or unavailable. Such counterfactuals should be based on a general deductive theory with clear microfoundations and scope conditions.[61] Elster, who also insists that good counterfactuals are derived from good theories, believes that there is only a narrow window for such experimentation: "The theory must be weak enough to admit the counterfactual assumption, and also strong enough to permit a clear-cut conclusion."[62]

Criteria that tie counterfactuals to established laws and statistical generalizations and attempt to limit second-order counterfactuals are superficially appealing. In practice, they are generally unworkable or would rule out some of the most important uses of counterfactual experimentation. The Fearon proximity criterion suffers from both defects. The requirement of a minimal cause that produces only a minimal effect is extraordinarily restrictive. Weber rightly observes that

[58] See M. Maruyama, "The Second Cybernetics: Deviation-Amplifying Mutual Causal Processes," *American Scientist* 51, no. 2 (1963). Maruyama contends that the very essence of the "butterfly effect" is that it cannot be discovered. He offers the example of a city built on the American plains because the first white settler awoke on an especially beautiful morning and took it as a sign to put down his stakes here rather than somewhere else. The coincidence of weather and intent was unpredictable and would be invisible as a cause to researchers a century later attempting to discover why the city grew up where it did.

[59] Fearon (fn. 29), 50.

[60] Dawes, "Counterfactual Inferences as Instances of Statistical Inferences," in Tetlock and Belkin (fn. 7)

[61] Edgar Kiser and Margaret Levi, "Using Counterfactuals in Historical Analysis: Theories of Revolution," in Tetlock and Belkin (fn. 7).

[62] Elster (fn. 38), 184–85.

"it rewards the psychologically easy and comfortable task of generat counterfactuals close to the margins of existing theories. It predispo toward varying only the familiar variables, the ones that we think know are tied into causal paths that we feel we know well." It also pı supposes that we know what "minimal" really means and that we ha a rather complete understanding of the behavior in question and its like consequences. "If we knew this," Weber continues, "we would no long need counterfactuals."[63] Counterfactuals almost always have multipl consequences, and a counterfactual powerful enough to test a theory tha makes only one small change in reality is probably an oxymoron.

Elster, Kiser and Levi, and Tetlock and Belkin favor counterfactual that are derived deductively from good theories.[64] This may be possible in a data-rich and reductionist field like cognitive psychology. But as Breslauer contends, it is hardly a realistic standard for counterfactuals in history and most of the social sciences, where after fifty years of intensive behavioral research there are no "established" theoretical laws or generalizations.[65] In political science, where researchers do not agree on the meaning of "cause" or "fact," it is unlikely that they will ever agree about what constitutes a good theory and, by extension, a useful or valid counterfactual.[66]

In the absence of established theory, Tetlock and Belkin worry that competing schools of thought will invent "counterfactuals of convenience." Breslauer and Herrmann and Fischerkeller report that this was a common practice in Soviet studies during the cold war. The widespread appeal of the expansionist "theory" of Soviet foreign policy made American policymakers and scholars receptive to counterfactuals that did not meet the Tetlock-Belkin criteria. Tetlock and Belkin would eliminate this problem by insisting that counterfactuals "must stimulate testable propositions that hold up reasonably well against new data." Herrmann and Fischerkeller point out that even if good evidence had been available, neither the expansionist not the defensive "theories" of Soviet behavior were formulated in ways to make them falsifiable.[67]

[63] Weber, "Counterfactuals, Past and Future," in Tetlock and Belkin (fn. 7), 278, 272.

[64] Elster (fn. 38); Kiser and Levi (fn. 61); and Tetlock and Belkin (fn. 23) also believe this is a reasonable standard for judging the plausibility of counterfactuals; Tetlock and Belkin acknowledge that "we should expect disagreement about what counts as well-established theory in world politics" (p. 27).

[65] Breslauer (fn. 34), 73; Steven Bernstein et al., "G-d Gave Physics All the Easy Problems: Adapting Social Science to an Unpredictable World," *European Journal of International Relations* 6 (January 2000).

[66] Weber (fn. 63), 271.

[67] Breslauer (fn. 34); Richard K. Herrmann and Michael P. Fischerkeller, "Counterfactual Reasoning in Motivational Analysis: U.S. Policy toward Iran," in Tetlock and Belkin (fn. 7); Tetlock and Belkin (fn. 23), 27.

Dawes's criterion of well-established statistical regularities is open to the same criticism. Based on an examination of the American literature on Iran, Herrmann and Fischerkeller conclude that "too often in world politics what is taken as a base rate for a generalization about the motives of another country is too much an ideological conviction and too little a product of deductive and empirical behavioral science."[68] Lebow and Stein have documented the same phenomenon with respect to deterrence; data sets used to test the strategy of deterrence were patently ideological in the cases they recognized as deterrence encounters and coded as successes for the West.[69]

There are other fundamental problems with attempts to derive counterfactuals from theory. Most theories rely on "structural" variables for their analytical power. To introduce variation in structure, it is necessary to make major changes in reality, changes that violate not only Fearon's proximity criterion but also any conception of a "minimal" or plausible rewrite of history. In international relations, most of our theories are at the levels of either the state (for example, democratic peace) or the system (for example, balance of power, power transition). To manipulate the relevant variables counterfactually, it would be necessary to make structural changes in the constitutions or capabilities of states or in the polarity of the system. It is inconceivable that any change of this kind could be brought about by a plausible-world counterfactual, unless it is introduced at some temporal remove and expected to have a long-term, cumulative effect. Multiple-step counterfactuals with unknowable second-order counterfactuals are unacceptable to theory-driven researchers. To change governments, capabilities, or polarity we must rely on miracle counterfactuals.[70] They require additional counterfactuals to bring them about, counterfactuals that researchers cannot in good conscience ignore because of their possible implications for the consequent. If we accelerated the decline of Great Britain in the nineteenth century relative to Germany to evaluate power-transition theory or if we made post-1918 Russia a democracy to test the relative importance of capabilities versus regime type, we would need to ask what could cause Britain's decline or Russia's emergence as a democracy. We would probably have to transform the economic development, class structure, and ideology of these states, and perhaps the outcomes of

[68] Herrmann and Fischerkeller (fn. 67), 146–47.

[69] Richard Ned Lebow and Janice Gross Stein, "Deterrence: The Elusive Dependent Variable," *World Politics* 42 (April 1990); and idem, *When Does Deterrence Succeed and How Do We Know?* (Ottawa: Canadian Institute for International Peace and Security, 1990).

[70] Fearon (fn. 29) acknowledges the utility of miracle counterfactuals but not the contradiction between their use and his requirement that counterfactuals be deduced from theories (pp. 60–65).

wars in which they had engaged. Any of these changes could signi cantly alter their foreign policies and perhaps the rules and norms the international system.

None of the miracle counterfactuals we could use to alter regim type, capability, or polarity can be derived from existing internation relations theories. They would all be ad hoc. The same holds true fo most plausible-world counterfactuals. Assassinations, ill-health, last minute changes in travel plans, undelivered messages, misunderstand ings, accidents, and coincidences—all the most credible vehicles o plausible-world counterfactuals—are random events, often unconnected with politics and certainly outside of any theory of politics. Why should this be considered a methodological weakness? In the health and physical sciences, where counterfactual experimentation is routine, counterfactuals are almost always independent of theory. They take the form of imaginary, often knowingly unrealistic parameters for key variables. Epidemiologists have robust equations to describe the spread of infectious diseases. In studying HIV, they might assign a range of arbitrary values to its virulence, the time that elapses between infection and the onset of symptoms, the rate of infection (by altering sexual practices or illicit drug use), or the nature and success rate of treatments, all with the goal of determining the combinations of social and medical practices that will be most effective in limiting the spread of the virus. So too in social science, the appropriate criteria for good counterfactuals have nothing to do with their origins, but rather concern how they are used.

This brings us to the nub of the problem. The scholars most troubled by the inherent unpredictability of counterfactual outcomes are those like Elster, Fearon, and Dawes who want to use counterfactuals to test propositions and theories. The less demonstrable the consequent, the less useful a counterfactual is for this purpose. Fearon is willing to consider a fallback position. Although it may be impossible to prove that x caused y, it may be possible, he suggests, to demonstrate that without x, whatever happened would not have been y.[71] The author employed this strategy to critique structural theories that attempt to explain transformations of the international system. I used counterfactual arguments to show that these transformations were contingent in a double sense. They were the result of independent chains of causation that produced nonlinear effects when they came into confluence. The transformations were also dependent on catalysts, whose presence was problematic and whose causes were independent of any of the underly-

[71] Fearon (fn. 29), 54.

ing causes of the transformations. Both the confluence and the catalysts could easily have been prevented by minimal rewrite counterfactuals. At best, therefore, structural theories can identify conditions under which transformations become possible.[72]

Scientific validation of a counterfactual is important only if the goal is to test a proposition or theory. Testing is only one part of theory building and by no means the most important part. Even within the neopositivist tradition, the prior steps of identifying important questions or anomalies and formulating theories to explain or resolve them are generally recognized as critical. I argued earlier that counterfactuals serve these ends admirably by making scholars more sensitive to contingency, helping them work through the implications of existing theories and identify gaps and inconsistencies in them. Theory building is only one goal of social science. It also aims to broaden our intellectual horizons and to provide methods relevant to assessing the relative benefits and value of policy outcomes, real and hypothetical. Counterfactual experimentation is essential to these tasks and can be used effectively without requiring certainty that antecedents will lead to specific consequents. Steven Weber has rightly observed that counterfactuals are better used as "mind-set changers" and "learning devices" rather than as data points in explanation."[73] We should worry less about the uncertainty of counterfactual experimentation and think more about its mind-opening implications.

For most of these purposes described above, the clarity, completeness, and logical consistency of the arguments linking antecedent to consequents are more important than their external validity. I accordingly propose eight criteria for plausible-world counterfactuals. Numbers 1, 2, 4, and 5 are drawn from or are variants of the Tetlock-Belkin list, number 6 was recently proposed by Tetlock, and numbers 3, 7 and 8 are the author's.[74]

1. *Clarity.* All causal arguments should define as unambiguously as possible what is to be explained (the consequent in counterfactual arguments), what accounts for this outcome (the antecedent), and the principle(s) linking the two. Good counterfactuals should also specify the conditions that would have to be present for the counterfactual to occur. Some historians have argued that timely public health measures could have significantly reduced the mortality in Europe associated with the Black Death pandemic of the fourteenth century. For Euro-

[72] Lebow (fn. 37).
[73] Weber (fn. 63), 270.
[74] Tetlock and Belkin (fn. 23), 16–32; Tetlock (fn. 56); Lebow and Stein (fn. 8), 146–47.

pean communities to have implemented these measures, they would have had to recognize that human intervention could affect the spread of disease, and they would have needed the authority and will to impose draconian measures on travel and trade over the likely objections of the wealthy and merchant classes.[75] Both additional conditions are unrealistic given the values, knowledge, and political structure of the age; large-scale quarantines would not be implemented to combat plague until the eighteenth century.[76] Plausible-world counterfactuals not only require realistic antecedents, but the antecedents themselves must not require other, implausible conditions or counterfactuals.

2. *Logical consistency or cotenability.* Every counterfactual is a shorthand statement of a more complex argument that generally requires a set of connecting conditions or principles. The hypothetical antecedent should not undercut any of the principles linking it to the consequent. A case in point is Robert Fogel's famous argument that if railroads had not existed, the American economy in the nineteenth century would have grown only slightly more slowly than it actually did because a strong incentive would have existed to invent the internal combustion engine sooner.[77] Elster has rightly objected that if the technology were present to invent and produce automobiles, it would almost certainly have also led to the development of railroads.[78]

3. *Enabling counterfactuals should not undercut the antecedent.* Counterfactuals may require other counterfactuals to make them possible (for example, for Richard Nixon to have been president at the time of the Cuban missile crisis, he would have had to have won the 1960 election, and that would have required significant changes in the political context at home and possibly abroad. These changes might have had significant implications for both American and Soviet foreign policy.) Researchers need to specify all important enabling counterfactuals and consider their implications for the consequent.

4. *Historical consistency.* Max Weber insisted that plausible counterfactuals should make as few historical changes as possible on the grounds that the more we disturb the values, goals and contexts in which actors operate, the less predictable their behavior becomes.[79] Counterfactual arguments that make a credible case for a dramatically

[75] Hawthorn (fn. 38), 31–60.

[76] William H. McNeill, *Plagues and People* (Garden City, N.Y.: Doubleday, 1976), 239–40, 254.

[77] Fogel, *Railroads and American Economic Growth: Essays in Econometric History* (Baltimore: Johns Hopkins University Press, 1964).

[78] Elster (fn. 38), 204–8; Tetlock and Belkin (fn. 23), 22–23.

[79] Weber (fn. 48).

different future on the basis of one small change in reality are very powerful, and the minimal rewrite rule should be followed whenever possible. The *nature* of the changes made by the experiment are nevertheless more important than the *number* of changes. A minimal rewrite that makes only one alteration in reality may not qualify as a plausible-world counterfactual if the counterfactual is unrealistic or if numerous subsequent counterfactual steps are necessary to reach the hypothesized consequent. A counterfactual based on several small changes, all of them realistic, may be more plausible, especially if they lead more directly to the consequent.

5. *Theoretical consistency.* There are few, if any, generally accepted theories in the social sciences, and none in international relations, comparative politics, or history. For purposes of counterfactual analysis, it is nevertheless useful to reference any theories, empirical findings, historical interpretations, or assumptions on which the causal principles or connecting arguments are based. This will provide readers with a more explicit perspective from which to evaluate the counterfactual's plausibility.

6. *Avoid the conjunction fallacy.* There are good statistical grounds for the "minimal rewrite" rule, as the probability of a consequent is the multiple of the probability of each counterfactual step linking the antecedent to it. We nevertheless need to recognize the conservative bias inherent in statistical reasoning. The laws of statistics indicate that the probability of *any* compound counterfactual is exceedingly low. This does not mean that the current state of affairs was overdetermined, only that it is very unlikely that hypothesized antecedents will produce *specific* consequences at any temporal distance. Social and political developments are highly contingent, and the future is undetermined—as was the past before it became the present. The long-term consequences of change are unpredictable. If Mozart had lived to sixty-five, today's world could well have turned out to be strikingly different from the world we know. But many alternative worlds are possible, and the probability of any one of them coming to pass is exceedingly low. Counterfactuals might have changed the world but in ways that become exponentially more difficult to track over time because of the additional branching points that enter the picture. As the probabilities associated with these outcomes will vary enormously, researchers accordingly need to specify if their counterfactuals are intended to produce a specific world, a set of worlds with particular characteristics, or *any* world (on a specific dimension) other than the one that actually came to pass.

7. *Recognize the interconnectedness of causes and outcomes.* Surgical counterfactuals are unrealistic because causes are interdependent and have important interaction effects. Changes we make in the past may require other changes to make them possible and may also produce additional changes beyond those we intend to lead to the consequent. History is like a spring mattress: if one of the springs is cut or simply subjected to extra pressure, the others will also to varying degrees shift their location and tension.[80] Earlier we considered the counterfactual that "President" Nixon would have ordered an air strike and follow-up invasion of Cuba, which in turn would have triggered a nuclear war. But if Nixon had been president, he probably would have committed American forces to save the faltering Bay of Pigs invasion and Castro would have been overthrown. A subsequent Soviet missile deployment would have become moot. Good counterfactuals must specify what else might change as a result of a hypothesized antecedent, and they must consider how the most important of these changes might interact and influence the probability of the consequent.

8. *Consider second-order counterfactuals.* Even when there is good to reason to believe that the antecedent will produce the desired consequent, the possibility remains that subsequent developments will return history to the course from which it was initially diverted by the antecedent. This might be the long-term result of enabling counterfactuals necessary to bring about the antecedent, of follow-on counterfactuals produced by the antecedent, of counterfactuals arising from the consequent, or of interaction among any combination of these counterfactuals. Interaction effects among second-order counterfactuals might be considered third-order counterfactuals, and they too can have profound consequences for the subsequent course of development.

Attempts to identify and analyze *all* of the counterfactuals arising from the antecedent and consequent would quickly lead to an infinite regress. Researchers should nevertheless try to identify what in their view is the most likely course of events that could unravel their consequent or negate its value as an outcome. The last point entails the recognition that we choose a consequent because of some larger effect it is intended to have. If other developments make it unlikely that the consequent will have that effect, it may lose its attractiveness. No counterfactual argument is complete without some argument about "alter-

[80] Friedrich Engels suggested something similar. History was a "parallelogram of forces." If one person shook his arms to move one corner of the parallelogram, it affected parts of the figure far away and far removed from intentions of the actor. Richard J. Evans, *In Defense of History* (New York: Norton, 1999), 118.

native" alternative futures and some assessment of their likelihood and implications for both the consequent and its value as a consequent.

These criteria will not allow researchers to validate plausible-world counterfactuals, but they will help them weed out poor counterfactuals primarily on the basis of clarity and logical and substantive completeness. Most of the criteria are not applicable to miracle-world counterfactuals, which, by definition, are not required to meet any real-world tests. The value of such a counterfactual is based entirely on its ability to provoke or, better yet, to compel researchers to think about issues and problems they would not otherwise address, or to look at them in a new light. For a field where careful, technical work is increasingly valued over imagination, miracle-world counterfactuals can refocus our attention on important, big questions.

UNDERSTANDING CHINA'S REFORM
Looking beyond Neoclassical Explanations

By SHU-YUN MA

Jane Duckett. *The Entrepreneurial State in China: Real Estate and Commerce Departments in Reform Era Tianjin*. London: Routledge, 1998, 273 pp.
Lance L. P. Gore. *Market Communism: The Institutional Foundation of China's Post-Mao Hyper-Growth*. Hong Kong: Oxford University Press, 1998, 354 pp.
Jean C. Oi. *Rural China Takes Off: Institutional Foundations of Economic Reform*. Berkeley: University of California Press, 1999, 253 pp.
Edward S. Steinfeld. *Forging Reform in China: The Fate of State-Owned Industry*. Cambridge: Cambridge University Press, 1998, 300 pp.

THEORETICAL CONTEXT: LOCAL STATE CORPORATISM VERSUS NEOCLASSICAL POLITICAL ECONOMY

POST-Mao reform in China has not only transformed the political, economic, and social dynamics of the country but also has provoked rethinking on a wide range of theoretical issues: How could China have achieved rapid economic growth without privatization? Why have Chinese officials not been resistant to market reform? What makes the Chinese state developmental rather than predatory? These questions pose important challenges to neoclassical political economy (NPE), which contends that economic growth is impossible under public ownership, that state bureaucrats always oppose marketization, and that states tend to be predatory.

Jean Oi was among the first to respond to those questions. In an article published in *World Politics* in 1992, Oi identified "local state corporatism" (LSC) as the secret of China's success. Her thesis has attracted wide attention among China scholars and has served as the common starting point of a number of studies on China's reform, including the recent books by Jane Duckett, Lance Gore, and Edward Steinfeld. More recently, Oi updated her arguments in a book-length study. The

World Politics 52 (July 2000), 586–603

urpose of this article is to show how these works attempt, and all fail, o establish China as a case to refute NPE. But first, a review of the genral arguments in the LSC-NPE debate may be helpful.

According to Oi, the decollectivization and the institution of the ousehold responsibility system in rural China transferred the income ights over agricultural production from collectives to individual houseolds. While this significantly enhanced the production incentives of easants, the change deprived local governments of a major source of ncome. At the same time, China's fiscal reform granted local governnents the right to retain part of the extra tax revenue they raised. That s, the higher the economic growth rate, the higher the tax revenue, and he greater the income of local governments. Given such a stake in ecoomic growth, local governments were motivated to mobilize and cordinate resources under their jurisdiction to engage in entrepreneurial ndeavors. They established and ran rural enterprises and took the rofits to pay for expenditures and reinvestment. In such a way, local overnments functioned like a large corporation with diversified busiesses, thereby serving as the engine of China's economic development.[1]

In Oi's view, her LSC thesis explains why China could achieve rapid conomic growth without privatization, why Chinese officials have not een resistant to the reform, and why the Chinese state has been develpmental rather than predatory. These phenomena indeed represent mportant challenges to NPE. Recalling Adam Smith, NPE starts with he view that individuals are ultimately self-interest seekers. They will ngage in economically productive activities only if they have maxinum control over the value they produce. Thus, ownership of private roperty is a prerequisite for economic prosperity. Applied to the case f China, NPE holds that privatization is a necessary, though not suffiient, condition for the country to make the transition from a central ommand system to an efficient market economy. The World Bank, the nost influential institutional advocate of such a view, thus contended hat "the [Chinese] government should *completely withdraw from* inherntly competitively structured industries where small and medium ized firms predominate."[2] While China's nonstate enterprises have een able to grow rapidly for almost two decades without the formal rivatization of state enterprises, "the limits of this strategy may be ap-

[1] Jean C. Oi, "Fiscal Reform and the Economic Foundations of Local State Corporatism in China," *orld Politics* 45 (October 1992).

[2] The World Bank, *China's Management of Enterprise Assets: The State as Shareholder* (Washington, .C.: World Bank, 1997), xiii, italics in original.

proaching," the World Bank said.[3] Such an NPE view has been echoed by scholars both in the West[4] and China.[5]

Another major proposition of NPE is that state bureaucrats, like individuals in the market, are also self-interest maximizers. Only different institutional environments give rise to different behaviors between market individuals and state bureaucrats. Given hard budget constraints, individuals in a market economy based on private property rights are conditioned to be self-utility maximizing, resulting in a Pareto-efficient distribution of resources. In contrast, state bureaucrats could achieve an "authoritative allocation of values" by exercising political power. Given such privilege, they are constantly tempted to extract resources from the economy using nonmarket measures. To protect vested interests, bureaucrats will resist any reform that aims at withdrawing political influence in the market. States thus tend to be predatory rather than developmental.

Using the above perspective, Yves Chevrier has shown how China's system of factory director responsibility failed due to resistance from vested interests of the party. The difficulty with China's reform is that "the existing power structure would have to find enough internal strength to reform itself while reforming the economy and society."[6] More specifically, Nigel Harris and David Lookwood argue that since communist states emphasized national defense, communist vested interests tend to concentrate in military industries. Hence, as in Russia, Vietnam, and Ukraine, the military industrial complex in China has been particularly resistant to privatization. As a result, the Chinese effort to transform the "war-making state" into a "market-facilitating state" has instead resulted in the creation of a "rent-seeking state."[7]

DUCKETT: CHINA AS AN ENTREPRENEURIAL STATE

As mentioned, the LSC thesis proposed in Oi's 1992 article was the first major attempt to show how China has proved an exception to NPE

[3] The World Bank, *Bureaucrats in Business: The Economics and Politics of Government Ownership* (Washington, D.C.: World Bank, 1995), 75.

[4] Wing Thye Woo, "The Art of Reforming Centrally Planned Economies: Comparing China, Poland, and Russia," *Journal of Comparative Economics* 18 (June 1994).

[5] Shaoguang Wang, "Learning by Debating: The Changing Role of the State in China's Economy and Economic Theories," *Policy Studies Journal* 23 (1995), 17.

[6] Yves Chevrier, "Micropolitics and the Factory Director Responsibility System, 1984–1987," in Deborah Davis and Ezra F. Vogel, eds., *Chinese Society on the Eve of Tiananmen: The Impact of Reform* (Cambridge, Mass.: Council of East Asian Studies, Harvard University, 1990), 131.

[7] Nigel Harris and David Lockwood, "The War-Making State and Privatization," *Journal of Development Studies* 33 (June 1997).

analysis. Duckett's *The Entrepreneurial State in China* is a further exploration of this issue. According to Duckett, post-Mao China is an "entrepreneurial state" (ES) characterized by the following major features: (1) the basic actors are individual state bureaus and their subordinate agencies; (2) these bureaus are directly involved in business; (3) they establish new profit-seeking businesses; (4) the new businesses are also risk-taking; (5) the state is adaptive to market reforms; and (6) ES activities are potentially productive and are thus not necessarily rent-seeking (pp. 14–15, 155–60).

While Oi focused on governments at the village, township, and county levels, Duckett narrows the unit of analysis to state bureaus. According to Duckett:

> The ES model focuses not on local government as a unitary actor, but on its constituent departments. It examines and describes the separate and autonomous profit-seeking economic activities of individual departments, explaining them as undertaken for the benefit of those individual departments, rather than for the development of the locality as a whole." (p. 13)

With the engagement of state bureaus in profit-seeking and risk-taking activities, Duckett argues, China has emerged into an ES rather than a predatory state. She believes that this has invalidated NPE's analysis (pp. 5–11). But from the viewpoint of NPE, Duckett's focus on the activities of bureaus does not provide sufficient basis for microanalysis. According to NPE, the basic unit of all social analyses should be individuals who share the same utility-maximizing natural desire. Human behavior must be explained in terms of the payoffs to the individuals involved rather than to any group to which they belong. Duckett "accepts that policy implementation and other 'state' activities are in the hands of individual officials" (p. 10). Throughout her analysis, however, she focuses on state bureaus, not individual bureaucrats.

The focus of Duckett's ES model (bureaus) is thus wider than that of NPE (individual bureaucrats) but narrower than that of LSC (local governments). The question then is, Why do bureaus provide a more meaningful level of analysis than individual bureaucrats and local governments? While deviating from NPE's individualist approach, Duckett identifies, in passing and in just two paragraphs, some personal incentives for officials to become entrepreneurial: higher status, more income, a better working environment, and diversion from the usual routine (p. 166). Haven't these ultimate concerns changed individual bureaucrats from opponents to supporters of market reforms? As long as *personal* incentives matter, the NPE rather than the ES perspective is

more relevant. For NPE, the disaggregation of local governments into individual bureaus could be useful if the change in group size is shown to have an impact on the tendencies to free ride and on principal-agent relations. But Duckett makes no efforts along these lines.

Duckett claims that she "criticises NPE not for its individualist method, but for its neglect of the context in which officials operate." Her examination of the responses of state officials to a particular set of structural constraints is, she thinks, "able to show how officials may adapt to marketisation if they can benefit from the opportunities that reform generates. State officials need not only seek to maintain the *status quo* to preserve vested interests; in some contexts they may have an interest in promoting reform" (pp. 10–11). What are those "contexts"? According to Duckett, several institutional changes have caused the emergence of the Chinese ES. In the first place, state bureaus have been subjected to increasing financial burdens as their tasks grow and expenditures rise on the one hand, but revenue falls on the other. State bureaus thus established new businesses not only to generate income but also to provide career outlets for officials laid off as a result of the state's reorganization. At the same time, the emergence of a nascent market in a still state-controlled economy, the continuation of communist party control and the work-unit system, and changing social attitudes toward doing business have facilitated state bureaus' engagement in profit-seeking activities. (pp. 161–65)

Aren't the above institutional changes almost the same as those described by Oi in her LSC? Failing to show that the institutional constraints that apply to bureau officials are more important than those that apply to local government officials, Duckett makes no real improvement to the LSC analysis. Duckett contends that as ideal types the LSC and ES "refer to quite different distinct kinds of activity." The essence of LSC is "non-profit-seeking co-ordination by local government to promote balanced local development," whereas ES "involves the individual, profit-seeking business activities of state bureau" (pp. 12–13). While emphasizing such distinctions between LSC and ES, Duckett also notes that the two models might be compatible with each other: "Since state entrepreneurialism refers to only one dimension of state activity, it may exist within developmental local governments. Governments may co-ordinate the promotion of developmental and infrastructural projects while still allowing individual departments to run businesses" (p. 173). If so, then when should each model be applied? Duckett answers, "If state entrepreneurialism were a dominant

the term ES as a description. If entrepreneurialism were less developed than co-ordinating activities, then the [LSC] might be the better characterisation" (p. 174). Hence, whether bureaus or local governments serve as the basic unit of analysis becomes unimportant, as both bureaus and local governments can be entrepreneurial or developmental. Then why differentiate between ES and LSC?

In fact, the above distinction that Duckett makes between LSC and ES does not seem meaningful. Under LSC, local governments not only coordinate economic activities; they also own and run rural enterprises.[8] Duckett criticizes Oi for making "occasional reference to individual bureaux' [*sic*] activities within the local corporatist state, [making it] unclear just what 'local government' refers to" (p. 173). But for Oi, bureaus and other kinds of local government organizations are alike in the sense that they may all engage in business activities. It is thus unnecessary to make any specific distinction between them.

Gore: Cadres as Bureaucratic Entrepreneurs

Gore's *Market Communism* also quotes Oi's LSC thesis. He even argues that LSC exists not only in Chinese villages but also in cities. According to Gore, "In so far as the institutional mechanisms of local state corporatism is concerned, I am in full agreement with these authors [including Oi]. And as such, local state corporatism exists not only in rural communities but in cities as well" (p. 160). This notwithstanding, Gore differs from Oi in his basic unit of analysis. While Oi was concerned with the response of local governments to institutional changes, Gore focuses on the behavior of individual bureaucrats.

Gore asks the central question, "What has *motivated* the dominant economic actors in the Chinese economy to *generate* such a long, sustained boom" (p. 13)? The dominant economic actors in reform China, according to Gore, are "bureaucratic entrepreneurs" (BEs). They are primarily "leading cadres" of the localities who are directly engaged in business activities aimed at promoting economic expansion (pp. 5, 96–97). By disaggregating local governments and bureaus into individual officials, Gore brings his analysis closer to the starting point of NPE than Oi or Duckett.

However, adopting a rather postmodernist stance, Gore argues that individual rationality is not inherited but is "socially constructed" (p. 28). That is, individual preference is defined by the institutional setting

[8] Oi (fn. 1), 115–22.

rather than by birth. Under private property rights, capitalist entrepreneurs react to "market profit alone" (p. 276). Reassignment of property rights, however, has little effect on the incentives of BEs. Noting that the earnings of private entrepreneurs were substantially higher than that of BEs, Gore writes: "If [BEs] are motivated by private financial gains alone, they would be better off to become private entrepreneurs" (p. 103).

In Gore's view, "private property rights play little or no role as an incentive" for BEs' activities, because BEs are "embedded in the basic institutions of Communism—the party-state hierarchy and its economic apparati." As such, their interests are intertwined with those of the communities their offices represent. Being simultaneous agents of the state and representatives of the community, they "pursue not only an economic rationale but also political and social objectives that come with their official status" (p. 101). Politically, BEs seek to advance within the hierarchy. Given bureaucratic competition and "the developmental pressure from above," they are induced to speed up local economic growth, as this would become an asset for political promotion. Socially, BEs try to satisfy as much as possible the social welfare demands within their organizations because of their commitment to their own work unit and the need for securing cooperation from their constituents. Only with continuous rapid economic growth can the rising demand for social welfare be met (pp. 106–13).

While playing down the importance of economic incentive, Gore notes that BEs do have economic concerns. However, like the above political and social factors, the economic interests of BEs are intertwined with those of their communities. "When they bargain for policy concessions, state subsidies and other resources on behalf of the community, they bargain for all the members including themselves; when they engage in 'group consumption', they are simultaneously engaged in private consumption" (p. 114). BEs often use their power and other communal resources under their control to start businesses. Yet, the purpose is not solely for private gain but "to generate additional income for their members" (p. 117).

Gore thus paints a picture in which there is no conflict between private and public interests. BEs strive for local development because it serves not only their personal but also their communal interests. Their personal welfare merges with the community. Gore argues that China has seen such results through its institutional setting, which is characterized by "a state-centred political economy" with continued public ownership and state administration of the economy on the one hand,

but weakening central control and emerging market forces on the other. "The mix feature[s] a unique combination of continued state-centricity and increased diversity and flexibility, and permits a market dynamism to emerge from amongst state hierarchies" (pp. 94–95).

In such a way, Gore claims that he provides an explanation of a phenomenon that NPE rejects as an accident—that a state can be developmental rather than predatory. He writes, "Proper institutional engineering can help to produce an incentive structure that may successfully contain the predatory tendencies of state elites. It may also be able to sustain a pattern of interaction among individual actors that can maintain the overall developmental orientation of the state." In Gore's view, his BE model "seems to be the only logical explanation for the puzzle of the developmental state" (p. 282).

Is this so? The answer depends on how we distinguish between developmental states and predatory states. If defined by outcome, there may be little question that China is a developmental state, as evidenced by the fact that the state-led reform has produced a two-decade growth. Yet, to define by outcome inevitably results in a tautology. An alternative is to define by a state's policies. Thus, a predatory state is one "whose policies can be shown to have benefited a monolithic self-seeking ruler and/or a bureaucratic elite," whereas a developmental state is one "that takes strong and direct measures to bring about economic development."[9] Conceptually such an approach seems more useful. But the question remains that if the strong and direct measures that bring about national development also benefit self-interest-maximizing bureaucrats at the same time, is the state developmental or predatory? This is exactly the difficulty that Gore created. When such an incentive system is in place, as Gore describes in his discussion of "bureaucratic entrepreneurialism," it can link bureaucrats' private interests with the national goal. Can we then still distinguish a developmental state from a predatory state? This is in fact a common methodological problem in the developmental state versus predatory state debate.

Perhaps more importantly, is the harmonization of private and public interests really a solution to this problem? Consider the case of a municipal economic bureaucrat in charge of a city's development plan in a liberal-capitalist state. The bureaucrat's private interests include personal income, career, and power. If the city prospers, he/she will reap rewards in all three categories—higher pay, greater chances of promo-

[9] Christer Gunnarsson and Mats Lundahl, "The Good, the Bad and the Wobbly: State Forms and Third World Economic Performance," in Lundahl and Benno J. Ndulu, eds., *New Directions in Development Economics* (London: Routledge, 1996), 257–58, 270.

tion, and increased authority on the job. That is, consistency between private interests and public interests, as depicted in Gore's BE model, could also exist in a liberal-capitalist system. Yet, this might not prevent the liberal-capitalist bureaucrat from placing private interests above public concern. We thus have no reason to expect the predatory tendencies of state bureaucrats under bureaucratic entrepreneurialism to be weaker than under liberal-capitalism. In Gore's view, predatory BEs may at the same time perform developmental functions. But so could their liberal-capitalist counterparts. The unique institutional structure in Gore's BE model simply produces no unique outcome.

Like Duckett, Gore also attempts to account for the apparently weak bureaucratic resistance to market reform in China. While Duckett explains this in terms of bureaus' interests in reform, Gore argues that "the rapid changes and the still-weak legal and procedural basis for coalition-building and other forms of effective political mobilizaton and political participation" in China have so far prevented the retrenchment of vested interests. As a result, bureaucratic resistance to reform "has been mostly passive and reactive, and has not been sustainable in the long run. Indicative of this is the fact that *the centre has eventually been able to push through all major reforms that it deems as essential, despite lengthy and repeated delays caused by local resistance*" (pp. 292–93, italics added). Here Gore makes no reference to his BE analysis. To the contrary, he contradicts his BE model by suggesting that central officials are developmental while local bureaucrats are predatory.

STEINFELD: CONSTRAINTS AS IMPETUS TO GROWTH

While Oi asked the question, How could China achieve economic growth without privatization? Steinfeld in *Forging Reform in China* asks just the opposite, Why has privatization in China not produced the economic efficiency that property-rights analysis would predict (p. 36)? Oi's question assumes that privatization has not taken place in China. But according to Steinfeld, "although Chinese reformers may not speak the language of privatization, they have certainly internalized the logic" (p. 34).

The reason for the apparent difference between Oi's and Steinfeld's theories is that Oi based her 1992 article on surveys of township and village enterprises (TVEs) conducted during the period 1986–91,[10] and

[10] Oi (fn. 1), 102.

Steinfeld, in his 1998 book, wrote about the more recent reform of large industrial state-owned enterprises (SOEs). When Oi carried out her study, privatization of TVEs was inhibited by political constraints. But when Steinfeld conducted his research, privatization of SOEs was taking place under the name of corporatization, a program that includes the evaluation of firm-level assets, issue of shares to private investors, and requirement of real returns on investment (p. 122). The two authors thus address different processes taking place in different sectors during different periods.

Yet, this does not mean that Oi and Steinfeld share no common ground. According to Steinfeld, privatization in China, as in Russia and Eastern Europe, has not brought about any significant improvement in economic efficiency. Rather, "in China today, the state firms that exhibit behavior most associated with private enterprise in the West—value maximization, strict budgeting, rational investment strategies, efficiency maximization—happen to be among the *least* privatized in the country. . . . Proximity to some forms of state control *is* affecting performance, but in the opposite way from what traditional property-rights theory would predict." In other words, like Oi, Steinfeld found evidence of growth *without* privatization in China. Steinfeld calls it "the great paradox of Chinese SOE reform" (p. 38).

The problem with applying property-rights analysis to transition economies, Steinfeld argues, is that two fundamental theoretical assumptions do not hold. First, for privatization to work, the principle of "rule of law" must be upheld so that "rights" can be made into really "enforceable claims." The socialist economies from which transition systems evolved, however, were ruled by commands, not law. What is required, therefore, is not privatization or the transfer of property rights but the creation of such rights from scratch. Since only the state has the legitimate authority to create rights, it is the solution to, as well as a cause of, economic inefficiency (pp. 38–40).

Steinfeld is of course right in emphasizing the importance of rule of law in the maintenance of private property rights. But the fact remains that some forms of private-property rights legislations have already been introduced in Russia, Eastern Europe, and more recently, China. While it is true that the rule-of-law condition was absent in former socialist countries, it may not be completely correct in today's transition economies. More importantly, it is not clear why the creation of private property rights and privatization should be treated as two separate strategies, as Steinfeld has done. Shouldn't the creation of private property rights be an indispensable component of privatization?

The second assumption of property-rights analysis Steinfield found problematic is that individuals are regarded as the basic economic actors. Given private ownership, rational economic beings are expected to maximize their personal utility, leading to Pareto-efficiency. "The problem is that transitional systems are industrialized economies. Because they are industrialized economies, the relevant economic actors are firms, and often extremely large firms at that." Steinfeld's unit of analysis (firms) is thus different from that of Oi (local governments), Duckett (bureaus), and Gore (individual officials). "Once firms are taken to be the major actors, however, the terms of debate shift dramatically. Property rights per se, particularly the degree to which those rights can be brought together and merged into the hands of clearly specified owners, recede in importance. Instead, a host of new issues take precedence" (p. 40). Steinfeld then goes on to show how the principal-agent problem characterizes modern industrial enterprises (pp. 40–44).

Transitional systems, as Steinfeld noted, are industrialized economies. But so are developed capitalist systems. Should we therefore also replace individuals with firms as the basic unit of analysis in capitalist economies? The principal-agent problem does illustrate the complexity arising from the separation of ownership from control; it shows that private property rights may not be a sufficient condition for efficient corporate governance. But to use it as an argument against the importance of privatization in transition economies is rather far-fetched. Finally, the principal-agent problem arises exactly because agents are found to be self-interest-seeking individuals rather than loyal servants of the firms they serve. Firms' behavior, therefore, is ultimately determined by individual rationality.

Dissatisfied with property-rights analysis, Steinfeld proposes what he calls a "constraints-based approach." In his view, the determining factor of enterprise behavior is the degree of hardness of budget constraint rather than ownership.

> For the specific case of Chinese SOES, an important variable to focus upon is access to soft subsidization—more specifically, access to the state banking system. Once firms have access to unlimited bailout loans, pressure to shift over to market-oriented behavior decreases for government regulators and enterprise managers alike. Opportunities for rent seeking are created, and ultimate constraints—namely, the bankruptcy and liquidation of the firm—never quite materialize. Predatory state taxation and managerial distortions of performance data can go on unabated. . . . In short, even as ownership is held constant, varying the firm's access to the banking system shapes the degree to which relevant economic actors engage in market-oriented behavior. (p. 47)

Hence, "*the imposition of hard budget constraints should be the prime goal of transitional reform*" (p. 247, italics in original).

According to Steinfeld, his constraints-based approach differs from property-rights analysis in that his approach views reform as a "clamping down," whereas the latter approach sees it as a "loosening up" (p. 9). Instead of assuming that economic dynamics will emerge once actors are granted certain property rights, Steinfeld holds that efficiency can be obtained only by imposing hard budget constraints on them. "Behavioral change is viewed not as a spontaneous response to new freedom but rather as a forced response to new pressures and strictures" (p. 229). Here Steinfeld mistakenly treats his constraints-based approach and property-rights analysis as two competing theories. Property-rights analysis does emphasize the importance of freeing economic actors to pursue personal interests but the general equilibrium is the condition that all individuals make optimal choices subject to hard budget constraints. It is the interactions between rights and constraints that result in Pareto-optimality.

For Steinfeld, Oi's notion of LSC is a property-rights analysis and is thus defective in its neglect of constraints. He challenges LSC by asking, Why are TVEs successful only in southern China but not in the inland or northern areas? While all local governments in the country are granted income rights over TVEs, why are there such regional differences in outcome? Applying his constraints-based approach, Steinfeld explains that due to the traditional secondary position of the south in China's command economy, state subsidies are less available in the south than in the inland and the north. It is thus the difference in the degree of hardness of budget constraints rather than ownership that accounts for the different economic performance of TVEs (pp. 237–41).

In her more recent work, Oi addresses this question of regional disparity without referring to Steinfied (see below). But to the extent that Steinfeld's criticism is valid, it is the problem of LSC, not of property-rights analysis. Moreover, if Oi committed the error of neglecting constraints, then Steinfeld incorrectly forgot the importance of rights. Property-rights analysis emphasizes both rights and constraints; rights must work with constraints and vice versa. It is Steinfeld who mistakenly separated the two.

Finally, while Duckett and Gore both attempt to show that states can be developmental, Steinfeld's constraints-based approach lends support to NPE's predatory state thesis. According to Steinfeld, "When the large firm is state-owned, predatory agencies extract profit in their

role as owners. If the firm is declared 'private,' those same agencies will extract tax money in their role as agents of governmental authority" (p. 62). That is the state will be predatory whenever possible, irrespective of ownership form.

Oi: Local State Corporatism Revisited

In *Rural China Takes Off,* Oi presents her latest version of LSC. While noting that the new tax system instituted in 1994 has reduced localities' claim to the residual, Oi believes that her LSC analysis is still valid. The reform has increased the center's share of tax revenue, but in absolute terms local governments have been guaranteed a minimum income not less than the 1993 level. More importantly, localities have retained their exclusive rights over a wide range of income, including sales tax, local-enterprise income tax, and individual income tax. Hence, in Oi's view, the incentives for practicing LSC have remained (pp. 54–56).

However, Oi is rather unresponsive to the works reviewed above, which all center on her LSC thesis. Duckett's ES analysis is used by Oi to supplement her point that local governments, given sufficient incentive and resources, can assume an entrepreneurial role. Oi notes that her study "recognizes that there are also distinct interests within local governments" (p. 193). But whether this necessitates narrowing the unit of analysis from local governments to bureaus, as Duckett does, is not discussed. Gore's study, which further disaggregates local governments and bureaus into individual bureaucrats, is not even included in Oi's bibliography.

While Oi focuses on local governments, her attitude to their basic components—individual officials—is ambiguous. She says that "there is no need to make assumptions about the nature of communist officials—either . . . they are corrupt or . . . they are motivated by lofty ideals that lead them to want to enrich their communities" (p. 6). Shortly thereafter, however, we find the following statement: "[This book] begins with the simple premise that local officials in a communist system, like officials in any political system, are rational actors who respond to incentives and existing constraints within the limits of their cognitive ability to evaluate alternatives and process information" (p. 7).

Indeed, only with such an NPE assumption do we find Oi's discussion of personal incentives relevant. According to Oi, fiscal contracting granted local governments autonomy in the use of residual they generated. "On a personal level

fers directly translated into lucrative (and legal) economic rewards for local officials. Local governments used the residual for bonuses. Linking cadre bonuses to the size of the residual gave local officials a direct stake in economic growth" (p. 49). Besides earning bonuses, individual officials who achieved outstanding economic growth in their localities were promoted to higher ranks within the bureaucratic hierarchy (p. 102). These personal incentives explain why Chinese officials were motivated to promote economic growth and not to obstruct market reform. They are the secret of the Chinese experience. As Oi notes, "China's economic reforms took off in many parts of the countryside precisely because the key entrepreneurs leading the charge were those who had the economic and political resources—the cadres themselves" (p. 192).

The above analysis should bring Oi in line with the NPE school, which emphasizes individuals as the basic unit of investigation. In fact, Oi claims that her study "is an institutional analysis of economic development at the local level that *views local officials as political economic actors* distinct from the central state" (p. 9, italics added). However, she does not adhere to such an approach. In most of her book Oi treats local governments rather than individual bureaucrats as the basic actors. For example, she writes, "[Even though the rights to residual] were granted to governments, *not individuals*, the re-allocation of property rights gave localities positive inducements to promote rapid economic growth" (p. 28, italics added). Here Oi tends to trivialize the importance of personal incentives.

The shift away from individual bureaucrats to local governments led Oi to the non-NPE view that privatization is not necessary. According to Oi, the success of the public-owned TVEs in China proved that growth can be achieved without privatization (p. 56). She claims that her study shows that "*individuals* need not have property rights over enterprise profits for economic growth to occur. Growth occurs as long as there are property rights for some organized entity. Local governments, if they have sufficient incentive and resources to pursue growth, can assume an entrepreneurial role" (p. 193, italics in original). Hence, in Oi's view, there is "no inherent reason why only individuals or privately held companies, as distinct from governments, can be entrepreneurs" (p. 10).

Oi overlooks her point that Chinese local governments can be distinguished from their counterparts in capitalist countries by their "corporatist" characteristics. According to Oi, the term LSC "was used primarily to highlight the workings of local government and the economy that it

firms as diversified corporations, redistributing profits and risks, a
thereby allowing the rapid growth of rural industry with limited r
sources" (p. 12). It is important to note that the corporations which (
compares with Chinese local governments are privately owned entitie
that is, the essence of LSC is that Chinese local governments are run lil
private corporations. Individual officials took bonuses from local gov
ernments' tax revenue, just as shareholders received dividends fron
joint stock companies. In the sense that public money (tax revenue) ha
been translated into bureaucrats' private property through such "corpo
ratization," privatization has actually taken place in China, however in-
formal it has been. Thus, Oi's assertion that China has achieved
economic success *without* privatization is not entirely correct.

Regarding Steinfeld's critique, Oi's response is rather enigmatic. She cited Steinfeld as an example of those who "have suggested that China's rural industry would be just where it is today, even without state intervention, simply as a result of market forces" (p. 4). Nowhere can we find in Steinfeld's work any such statement. What he did say is that the rural economic takeoff in China has not been a "national" phenomenon; rather, it has been restricted to "areas neglected during the command period," such as the southern provinces (p. 235). Steinfeld's point is not that market forces have prevailed in these neglected areas. Instead, he uses the regional differences to ask Oi the very valid question, If fiscal contracting is the cause of China's rural prosperity, then why has this national reform worked only in the south but not other areas? As mentioned above, utilizing his constraints-based approach, Steinfeld explains this in terms of the difference in the availability of soft credits among Chinese regions.

Without referring to Steinfeld's point, Oi acknowledges the problem of variation of growth as follows:

> All parts of China's countryside were subject to the incentives embedded in decollectivization and the fiscal reform, but only a portion of the countryside has successfully developed rural industry. Some areas have been left out of the overall trend of industrial growth and the prosperity that has come with it. *In those areas that have succeeded in industrializing*, various forms of ownership have been adopted. Some counties, townships, and villages have relied primarily on collectively owned enterprises, whereas others have favored private ownership. (p. 58, italics added)

Two questions arise from these statements. First, why have the nationwide economic incentives produced growth only in some but not all rural areas? Second, within those areas that have achieved industrialization, why have some adopted mainly collective ownership while oth-

s have more privately owned enterprises? Oi argues that the economic centives provided by fiscal contracting were intervened by the politi-l attitude toward private ownership. She devotes most of Chapter 3 show how the political factor has affected the form of ownership *in eas that have succeeded in industrializing.* That is, Oi addresses the sec-ıd question, although not necessarily satisfactorily, and does not even ckle the first (and perhaps more important) question. Steinfeld iden-ies this as the central methodological problem of LSC: "By examining ıly situations in which TVEs have thrived, LSC approaches commit the ror of selecting on the dependent variable, to use social science termi-ology. By trying to explain what they themselves define as a constant VE growth) and then by selecting only successful TVEs, LSC propo-nts have no way of proving that state intervention is the key" (p. 237).

Toward the end of her book, Oi writes: "In most villages, the deter-ining factor is the degree of local industrialization. Local govern-ents that control a primarily agricultural, particularly grain-based, conomy have few options other than to levy ad hoc surcharges and arious other fees and penalties. [LSC] is unlikely to exist in such vil-ges. The greater the level of industrialization, the more likely that the ıcal government will act in a corporate manner to intervene, extract, nd redistribute income" (p. 191). On this point Oi seems to be con-ısed about the direction of her explanation. Is it LSC that has facili-ıted rural industrialization or vice versa?

CONCLUSION

'he four works reviewed above enhance our knowledge of China's re-orm but they do not adequately establish China as a case to refute the PE perspective. Deviating from the NPE's axiomatic approach of treat-ıg single individuals as the basic political and economic actors, Oi, Duckett, and Steinfeld respectively focus on the behavior of local gov-rnments, bureaus, and firms. But they all fail to explain why their units f analysis, when compared with individual bureaucrats, are more elevant respondents to institutional changes. Gore alone focuses on in-ividual bureaucrats. Yet his argument that China's institutional setting an produce altruistic bureaucrats is unconvincing.

However, the failure of these searches for non-NPE explanations does ot imply that the NPE approach is adequate. In fact, so far NPE has not rovided a satisfactory answer to the China puzzles either. To generate ome kind of consensus on the issue, I would like to suggest the fol-owing method-

tions for future research. Methodologically, to solve China's economic "miracle," we must start with NPE's egoistic assumption because we have sufficient grounds to believe that Chinese bureaucrats have adopted such entrepreneurial behavior purely out of self-interest. There is ample evidence showing that Chinese bureaucrats make use of their public positions to reap private gains, and it is often hard to distinguish their "entrepreneurial" activities from corruption. It is exactly the existence of such bureaucrats that makes the Chinese case puzzling. The NPE perspective has the strength of identifying the China puzzles, and the challenge is to explain these puzzles without departing from NPE's individualistic starting point.

Empirically, we need further research to show how real the puzzles are. For example, a common measure must be established to determine the extent of privatization that has, or has not, taken place in China. This must be done before tackling Oi's question about how China could achieve economic success without privatization and Steinfeld's opposing question about why privatization in China has failed to produce economic efficiency. Likewise, before attempting to explain why Chinese bureaucrats have not been resistant to market reform, more detailed empirical work should be carried out to examine whether any geographical, sectoral, and generational variations in bureaucratic resistance exist. This will improve the focuses of our analyses. Finally, clear and operationalizable criteria must be identified to distinguish between developmental and predatory behaviors. As mentioned, this is particularly necessary, and also particularly difficult, in situations where the incentive system seems to be able to harmonize conflicts between personal and public interest.

Theoretically, insights can be sought from the ongoing collaboration between NPE and institutionalism. While NPE explains behavior in terms of utility maximization of egoistic individuals, institutionalism emphasizes the importance of institutional constraints. The complementarity of the two approaches has produced a fast-growing literature.[11] In fact, some applications of such joint perspectives have been made to the Chinese case,[12] but they do not tackle directly the puzzles that concern us. Efforts along these lines will involve the identification

[11] B. Guy Peters, *Institutional Theory in Political Science: The New "Institutionalism"* (London: Pinter, 1999), 43–62.

[12] Steven N. S. Cheung, *Will China Go 'Capitalist'?* (London: Institute of Economic Affairs, 1986); Steven L. Solnick, "The Breakdown of Hierarchies in the Soviet Union and China: A Neoinstitutional Perspective," *World Politics* 48 (January 1996); and David L. Wank, "The Institutional Process of Market Clientelism: Guanxi and Private Business in a South China City," *China Quarterly* 147 (September 1997).

f the institutional constraints that previous research may have nelected and an explanation of how Chinese bureaucrats have responded) these constraints in a way that makes the Chinese case so special. Iowever, the limitations of such an approach must be specified to avoid making nonfalsifiable conclusions.

Comparatively, it would be useful to situate the Chinese case within the larger perspective of transitions in other postcommunist countries. In particular, Russia would be a good case for comparison, as it shares with China a similar background of indigenous communism but diverges sharply in terms of reform performance. Some have argued that such a comparison is inappropriate since China has some unique favorable initial conditions, such as a deeper sense of crisis, a less centralized agricultural sector, greater availability of surplus labor, a shorter history of communism, higher saving habits, and stronger ties with its emigrant community.[13] However, others have contended that no comparison is perfect. Instead of passively enjoying the favorable setting, China has actively promoted some "right" policies, including adopting an incremental approach, prioritizing agricultural and light industrial reforms, and opening foreign trade and investment.[14] The larger questions are, What leads China to such "right" policy choices? Are they outcomes of specific institutions? If so, are these institutions transferable? After all, it is the transferability of the Chinese experience that has the most practical implications.

In conclusion, the Chinese reform has provoked intellectual rethinking of the NPE perspective. The four works reviewed here represent different attempts in this regard. None of them, however, offer a better alternative. The search for the secret of China's economic "miracle" must be continued.

[13] Anders Aslund, *How Russia Became a Market Economy* (Washington, D.C.: Brookings Institution, 1995); Yuan Zheng Cao, Gang Fan, and Wing Thye Woo, "Chinese Economic Reforms: Past Successes and Future Challenges," in Woo, Stephen Parker, and Jeffrey D. Sachs, eds., *Economies in Transition: Comparing Asia and Eastern Europe* (Cambridge, Mass.: MIT Press, 1997).

[14] Marshall I. Goldman, *Lost Opportunity: Why Economic Reforms in Russia Have Not Worked* (New York: W. W. Norton, 1996); and Barry Naughton, "China's Economic Success: Effective Reform Policies or Unique Conditions?" in Kazimierz Z. Poznanski, ed., *The Evolutionary Transition to Capitalism* (Boulder, Colo.: Westview Press, 1995).

INDEX TO VOLUME 52

WORLD POLITICS

A Quarterly Journal of International Relations

Volume 52

October 1999–July 2000

The editors invite submission of articles bearing upon problems in international relations and comparative politics. Manuscripts and notes should be double-spaced throughout and submitted in triplicate, together with an abstract. Word count should be indicated. Manuscripts may be up to 10,000 words, including notes; those that exceed the limit by more than 10 percent will *not* be considered. Address: WORLD POLITICS, Bendheim Hall, Princeton University, Princeton, NJ 08544.

Authors may expect to be notified of the editors' decision within four months. Procedures for reviewing manuscripts are based on the anonymity of the author and the confidentiality of readers' and editors' reports; hence, self-references should be removed. Manuscripts are refereed by the editors and by other scholars at Princeton and at other institutions. WORLD POLITICS does not accept manuscripts that have already been published, are scheduled for publication elsewhere, or have been simultaneously submitted to another journal; this applies to both print and online formats. Statements of fact and opinion appearing in the journal are made on the responsibility of the authors alone and do not imply the endorsement of the editors or publisher. The journal does not publish communications to the editor or rejoinders to specific articles; scholars who believe they have been challenged are encouraged to submit an article that will advance the scholarly debate. The WORLD POLITICS home page is found at http://www.wws.princeton.edu/world_politics

Reprint permission: Permissions Department, The Johns Hopkins University Press, 2715 N. Charles Street, Baltimore, MD 21218-4319.

Subscriptions: WORLD POLITICS, The Johns Hopkins University Press, Journals Publishing Division, 2715 N. Charles Street, Baltimore, MD 21218-4319. Phone: (410) 516-6987; FAX: (410) 516-6968; toll free (800) 548-1784. Individuals: $28 per year; institutions: $86 per year. Single copies: individuals, $8.00, institutions, $21.00. For postage and handling in Canada and Mexico, add $4.00 per year; outside of North America, $9.00.

Back Issues: WORLD POLITICS is available in microfilm or xerographic copies from University Microfilms, Ann Arbor, MI 48106. Two years after publication, complete volumes are available in microfiche form from KTO Microform, Route 100, Millwood, NY 10546. Individual issues in the current and two preceding volumes may be ordered from The Johns Hopkins University Press, Journals Publishing Division, 2715 N. Charles Street, Baltimore, MD 21218-4319.

Advertising: Journals Advertising Coordinator, The Johns Hopkins University Press, Journals Publishing Division, 2715 N. Charles Street, Baltimore, MD 21218-4319.

WORLD POLITICS is indexed by *ABC POL SCI, Book Review Index, Combined Retrospective Index Sets (CRIS), Combined Retrospective Index to Book Reviews in Scholarly Journals, 1886–1974 (RSJ), PAIS Bulletin Social Science Index,* and *Periodica Islamica;* it is abstracted and indexed in *United States Political Science Documents, International Bibliography of Periodical Literature (IBZ), International Bibliography of Book Reviews (IBR),* and *Current Military and Political Literature.* Abstracts of articles also appear in *Book Review Digest, Historical Abstracts, International Development Abstracts, International Political Science Abstracts, Political Science Abstracts,* and *Social Sciences Abstracts*

WORLD POLITICS (ISSN 0043-8871). Published quarterly by The Johns Hopkins University Press. Vol. 52, October 1999–July 2000. Periodicals postage paid at Baltimore, MD, and additional mailing offices. POSTMASTER: Send address changes to WORLD POLITICS, The Johns Hopkins University Press, Journals Publishing Division, 2715 N. Charles Street, Baltimore, MD 21218-4319. Printed in the United States of America by Capital City Press.

CONTENTS